Bioinformatika - Proteomika

Medzihradszky (Fölkl) Katalin
Adjunct Professor of Pharmaceutical Chemistry,
University of California San Francisco
és
Proteomikai Kutatócsoport, SzBK
Csoportvezető, tudományos főmunkatárs
e-mail: folkl@cgl.ucsf.edu vagy folkl@brc.hu

Darula Zsuzsanna
Proteomikai Kutatócsoport, SzBK, zdarula@brc.hu

Bioinformatika úgy általában

- adatbázisok felépítése
- ezen adathalmazok statisztikai analízise, matematikai értelmezése
- szabályszerűségek meghatározása
- szabályok hasznosítása – predikciók pl. funkcióra, rokonságra, eredetre, szerkezetre stb.

Bioinformatika úgy általában

- „összehordott” adatokból építkezünk
- többnyire elmélet eleinte
- a gyakorlati hasznosítás majd a végén jön
- persze ez visszahat az elméletre

Mit lehet kiaknázni?

- genomiális adatbázisok
- fehérje szekvencia adatbázisok
- fehérje 3D szerkezeti adatbázisok
- stb, stb

Meghatározó faktor:

Mennyire megbízhatóak az adatok?

PROTEOMIKA

A sejt fehérjetartalmának kvalitatív és kvantitatív jellemzése

Mi van jelen? Mennyi? Milyen formában?

Lokalizáció? Kölcsönhatások? Partnerek?

A proteomika különleges jellege

- nagyon is gyakorlati tudomány
- adatokat produkálunk, amit informatika segítségével adatbázisokból értelmezünk
- bővítjük/építjük az adatbázisokat, ill. új adatbázisokat hozunk létre
- új bioinformatikai eszközöket fejlesztünk ki az eredmények alapján

Problémák

- állandó dinamikus változások
- hatalmas mennyiségi különbségek
- igen eltérő fizikai tulajdonságok/ vagy csak árnyalatnyi különbségek – ugyanannak a génnek a termékeire

Még mindig „problémák”

Poszt-transzlációs módosítások

Idő-, faj-, állapot-, lokalizáció- stb. - függő

- permanens vs. dinamikus
- teljes vs. részleges
- jelentős vs. csekély méretváltozás
- hidrofil
- hidrofób

Nem árt érteni a biológiához!

Problémák

- Mikor és hogyan veszünk mintát?
- Hogyan tároljuk a mintákat?
- Hányszor ismételjük meg az analízist?
- Milyen statisztikai módszereket használunk?

Láthatóvá tenni a komplexitást...

Frakcionálni kell !

1D-, 2D-, multidimenziós módszerek

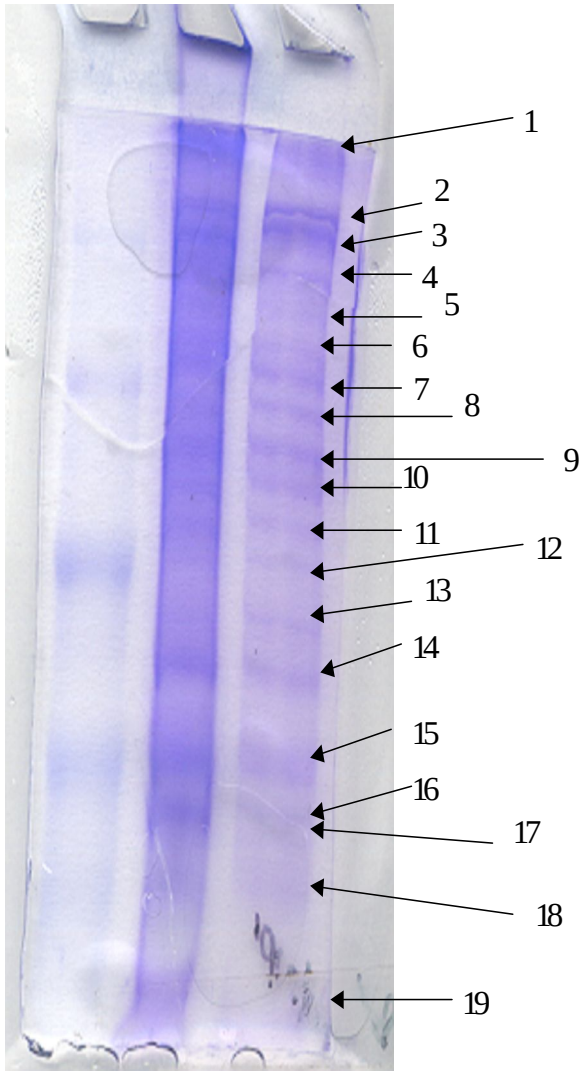
Mi mindent kell meggondolni?

- Mennyire univerzális a módszer?
- Felbontása?
- Dinamikus tartománya?
- Követhetősége?
- Reprodukálhatósága?
- Kvantitatív-e?

1D-SDS-PAGE

* Minden belemegy a gélbe, de kicsi a felbontás.

* Várhatóan fehérje-elegyeket kell analizálni.



2D-elfo

1. dimenzió: izoelektromos fókuszálás

→ pI szerint - pH 3-10

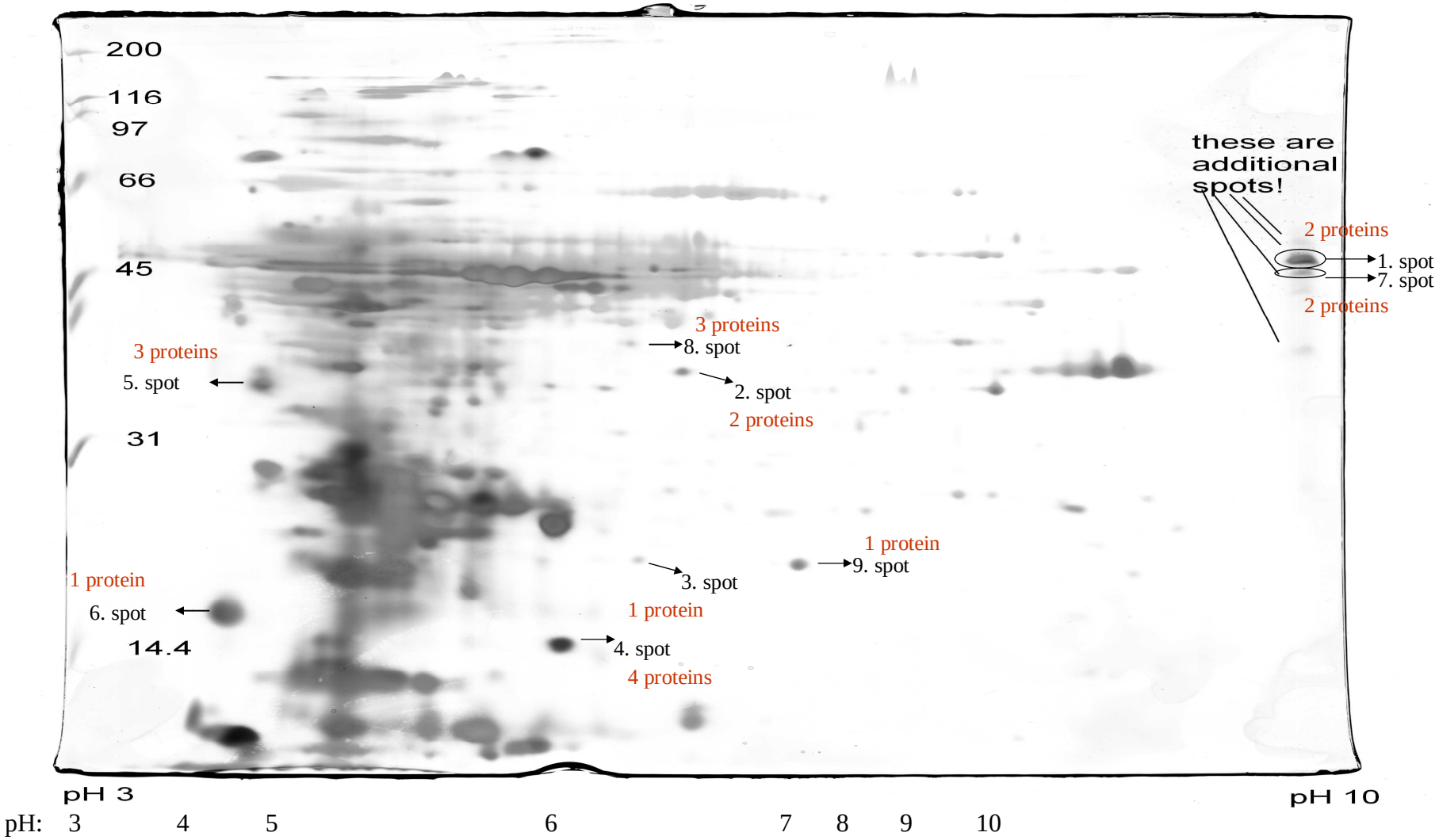
DE

savas vagy bázikus fehérjék NEM fókuszálódnak

membrán-fehérjék kicsapódnak

– ionos detergens nem használható

2. dimenzió: SDS-PAGE



Több példányban elkészítendő, statisztikai analízisre alkalmas

Vannak 2D-gél adatbázisok is

- http://ca.expasy.org/ch2d/2d-index.html

→**Multi species**

World-2DPAGE Portal;

SWISS-2DPAGE ;

SIENA-2DPAGE;

2-DE database at Max Planck Institute for Infection Biology;

2-DE maps at ISPA,

Speciálisabb adatok

- Toothprint (2-DE gels of dental tissues - University of Otago, Dunedin, New Zealand)

  RatEnamel

- JPLS Proteomics database at Ludwig Institute for Cancer Research, Melbourne, Australia

  Human Breast carcinoma cell line,
  Human Placenta,
  Human Colorectal cell line
  Mouse colonic crypts

- BALF 2D Database at Department of Biological Chemistry, University of Mons-Hainaut, Belgium

  Human, Mouse Broncho-alveolar lavage fluid

File Edit View History Bookmarks Tools Help

http://ca.expasy.org/swiss-2dpage/ | 2D gel database

Yahoo! Mail - folkl@y... | Save to My Web | my web | PP on lorenzetti | PP on cassatt

SWISS-2DPAGE | WORLD-2DPAGE List - 2-D PAGE data...

Maps

[experimental info]

[protein list]

[graphical interface]

Query Remote Interfaces:

World-2DPAGE

Exclude local DBs

has only effect if a remote interface is selected

Protein list for ARABIDOPSIS 79 identified spots / 57 protein entries

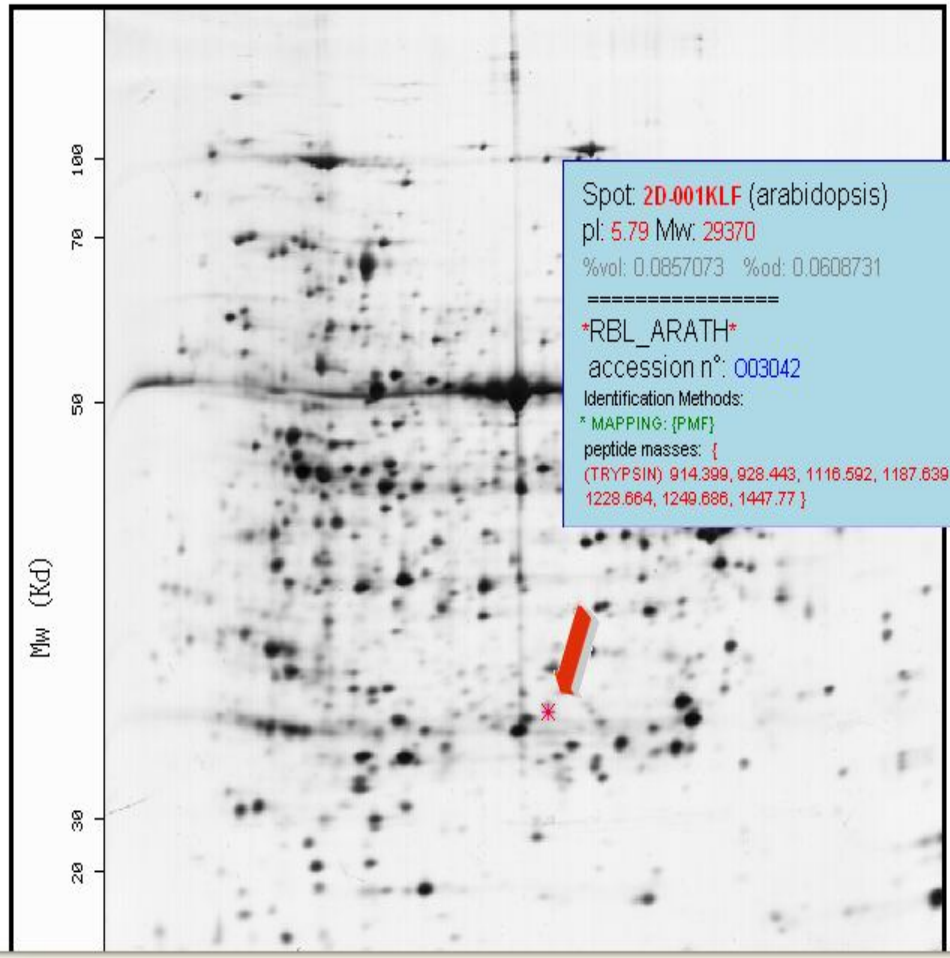
Gene Name	Description	Accession n°	Spot ID	Exp *pI*	Exp *Mw*	Mapping Methods	References
rbcL	Ribulose bisphosphate carboxylase large chain (EC 4.1.1.39) (RuBisCO large subunit)	O03042 (RBL_ARATH)	2D-001KKP	5.55	30'287	{PMF}	1
			2D-001KL6	5.79	29'705	{PMF}	1
			2D-001KJ2	5.92	32'050	{PMF}	1
			2D-001KLF	5.79	29'370	{PMF}	1
			2D-001K6O	5.71	50'864	{Gm, PMF}	2, 1
			2D-001K74	5.64	51'074	{PMF}	1
			2D-001K9S	6.15	47'034	{PMF}	1
			2D-001KBM	6.10	43'852	{PMF}	1
			2D-001KHU	6.02	33'897	{PMF}	1
			2D-001K7A	5.78	51'284	{PMF}	1
CRT1	Calreticulin-1	O04151 (CALR1_ARATH)	2D-001K5F	4.41	55'005	{Gm}	2, 1
OrderedLocusNames=At3g16470	Myrosinase-binding protein-like At3g16470	O04309 (MB31_ARATH)	2D-001KAA	5.22	46'361	{PMF}	2
OrderedLocusNames=At2g47470	Probable protein disulfide-isomerase A6 (EC 5.3.4.1) (P5)	O22263 (PDIA6_ARATH)	2D-001KEG	6.12	39'236	{PMF}	2
CIMS	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14) (Vitamin-B12-independent methionine synthase isozyme) (Cobalamin-independent methionine synthase isozyme)	O50008 (METE_ARATH)	2D-001K0N	6.15	76'088	{Gm}	2, 1
CPN21	20 kDa chaperonin, chloroplast (Protein Cpn21) (Chloroplast protein Cpn10) (Chloroplast chaperonin 10) (Ch-CPN10) (Chaperonin-20)	O65282 (CH10C_ARATH)	2D-001KMV	5.25	27'447	{PMF}	2

Done

http://ca.expasy.org - SWISS-2DPAGE - Mozilla Firefox

Micro-Sequencing / Tagging ○ Gel Matching ▫ Comigration ◦ Immunobloting

[estimated location]



Spot: 2D-001KLF (arabidopsis)
pI: 5.79 Mw: 29370
%vol: 0.0857073 %od: 0.0608731
==========
RBL_ARATH
accession n°: O03042
Identification Methods:
* MAPPING: {PMF}
peptide masses: {
(TRYPSIN) 914.399, 928.443, 1116.592, 1187.639, 1228.664, 1249.686, 1447.77 }

Done

SWISS-2DPAGE

Maps

[experimental info

[protein list]

[graphical interfa

Query Remote I

World-2DPAGE

Exclude loca

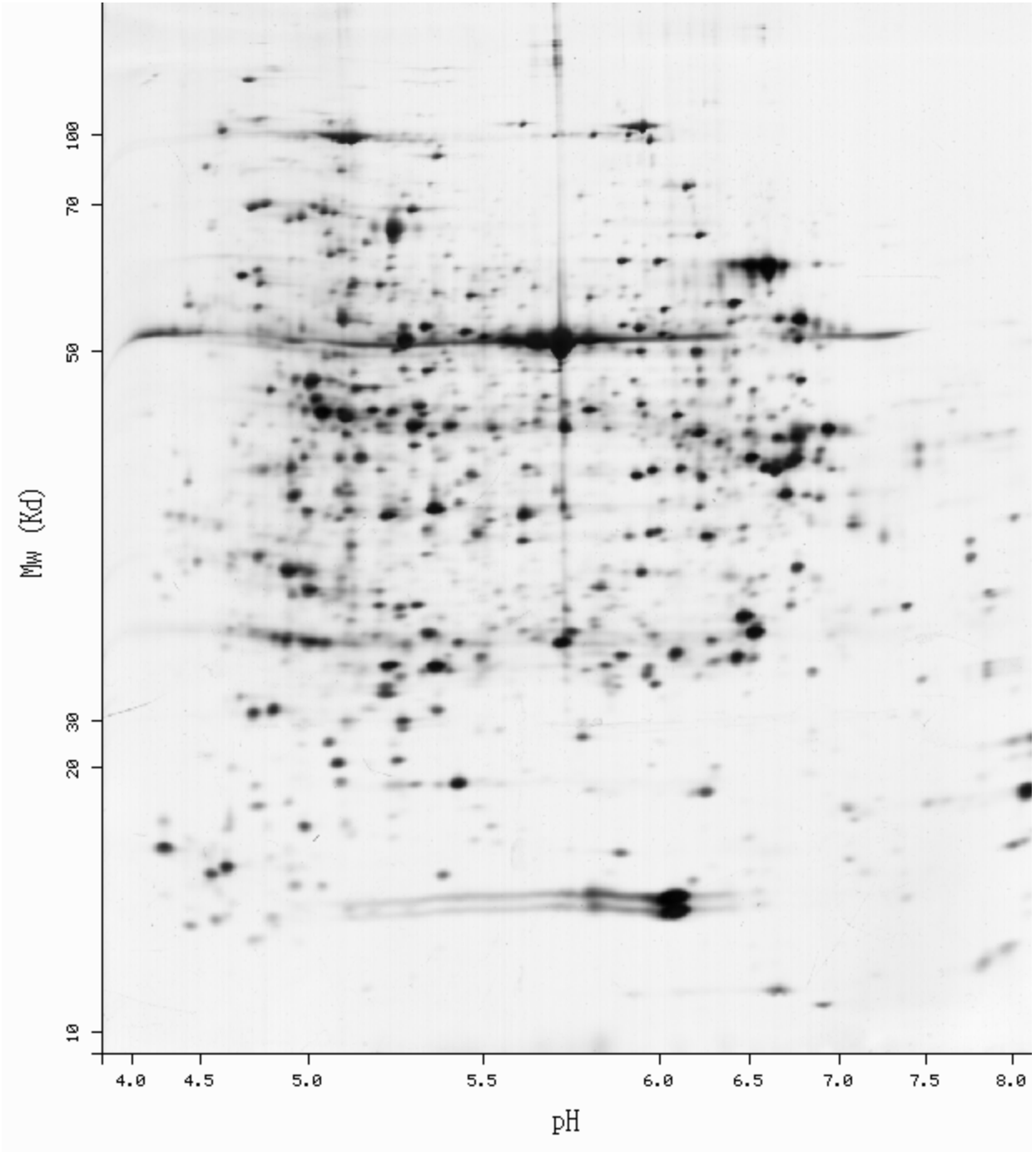
has only effect if a interface is selecte

ein entries

t ID	Exp pI	Exp Mw	Mapping Methods	References
1KKP	5.55	30'287	{PMF}	1
1KL6	5.79	29'705	{PMF}	1
1KJ2	5.92	32'050	{PMF}	1
1KLF	5.79	29'370	{PMF}	1
1K6O	5.71	50'864	{Gm, PMF}	2, 1
1K74	5.64	51'074	{PMF}	1
1K9S	6.15	47'034	{PMF}	1
KBM	6.10	43'852	{PMF}	1
1KHU	6.02	33'897	{PMF}	1
1K7A	5.78	51'284	{PMF}	1
1K5F	4.41	55'005	{Gm}	2, 1
1KAA	5.22	46'361	{PMF}	2
KEG	6.12	39'236	{PMF}	2
1K0N	6.15	76'088	{Gm}	2, 1
		synthase isozyme)		
CPN21	20 kDa chaperonin, chloroplast (Protein Cpn21) (Chloroplast protein Cpn10) (Chloroplast chaperonin	O65282 (CH10C_ARATH)	2D-001KMV	5.25	27'447	{PMF}	2

javascript:OpenViewerForSpot('2D-001KLF', 'ARABIDOPSIS', '1')

start | SWISS-2DPAGE - Moz... | http://ca.expasy.org... | BioInfo04 | Inbox - Thunderbird | Re: MS még egyszer ... | Document1 - Microsof... | EN | 6:58 AM


Mw (Kd)
100
70
50
30
20
10
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
pH

2D-elfo másképpen

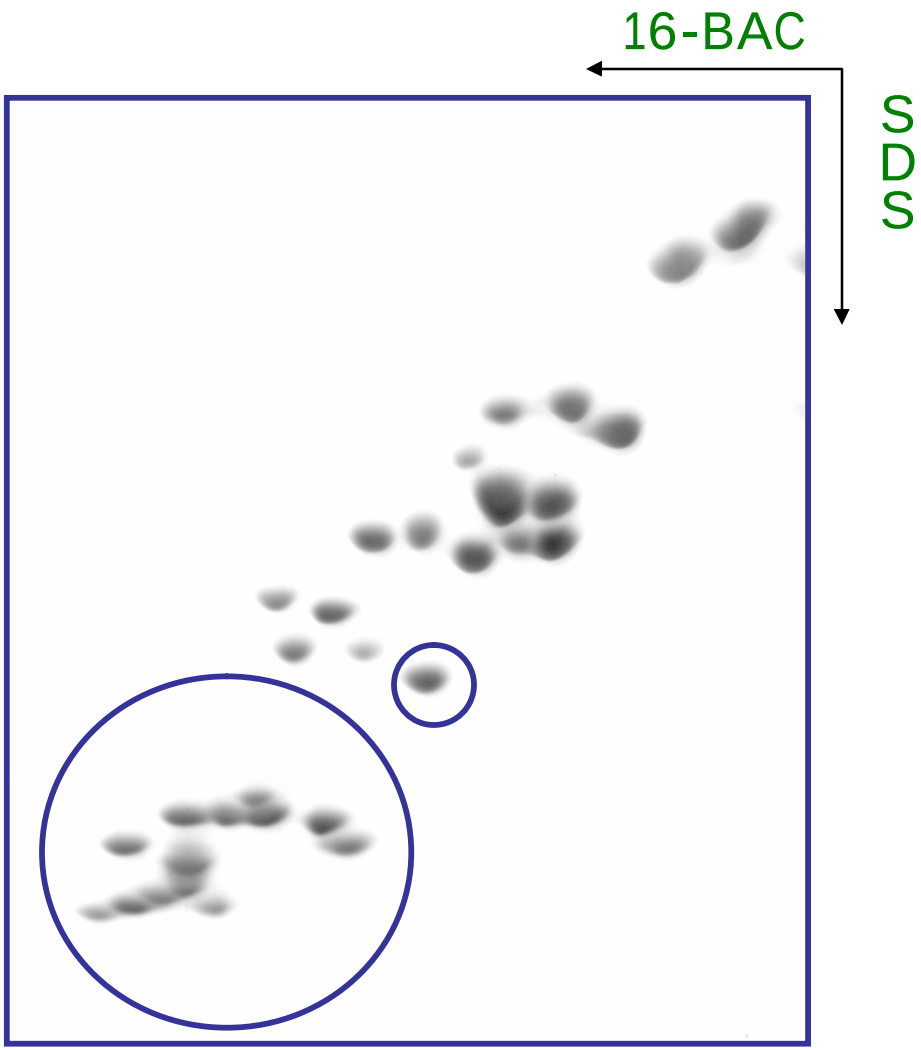
1. dimenzió: 16 BAC-PAGE

(pozitív „fejű” detergens)

2. dimenzió: SDS-PAGE

Minden belemegy a gélbe, de átlósan frakcionálódik – kis felbontás

Speciális 2D elválasztás



Pros35	10/14 (71%)	33%
Prosα7	34/36 (94%)	63%
Prosα6	12/12 (100%)	42%
Pros28.1	14/21 (66%)	54%
Pros29	19/23 (83%)	59%
Prosβ2	13/15 (86%)	42%
GC12000	13/14 (93%)	35%
ProsMA5	22/44 (50%)	56%
Pros25	12/15 (80%)	50%
Prosβ3	18/24 (75%)	53%
Pros26	6/15 (40%)	46%
Prosβ5	11/12 (92%)	26%
GC17331	11/23 (47%)	46%
l(2)05070	15/19 (79%)	72%

Festés

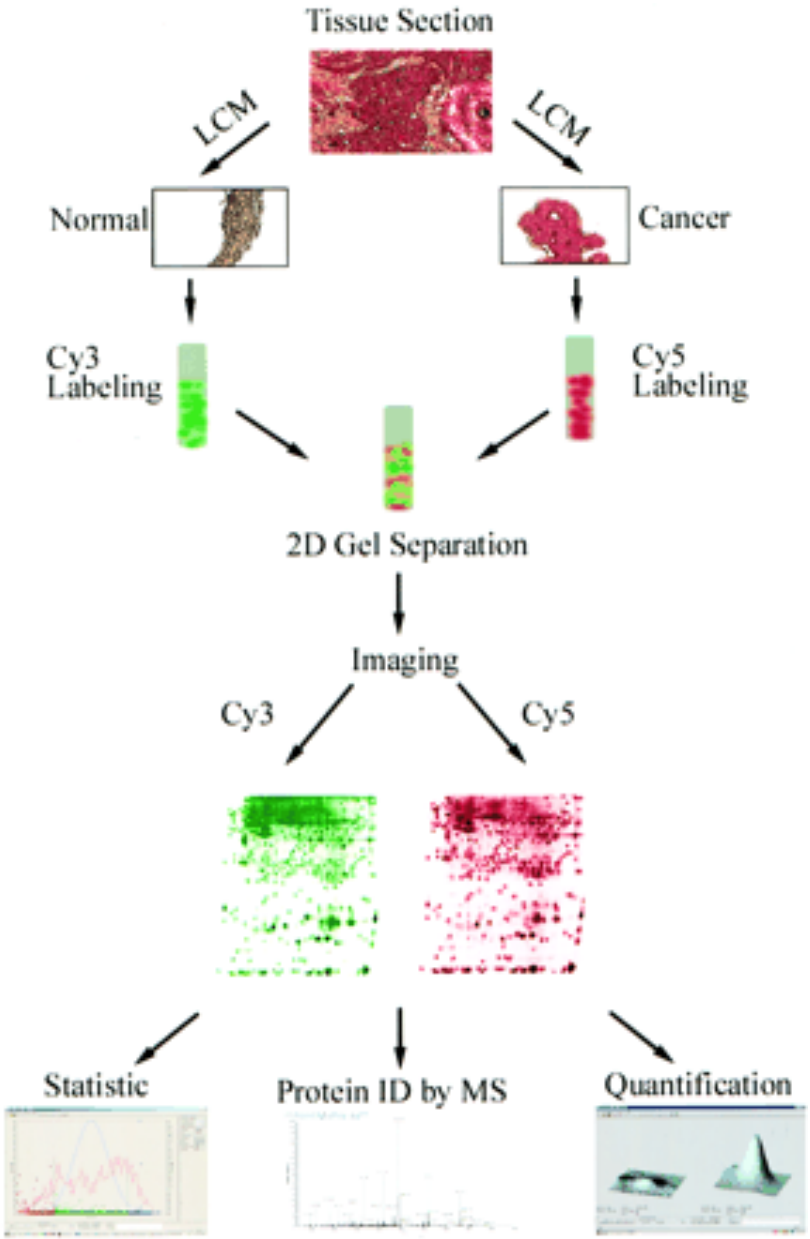
- **C**ommassie **B**rilliant **B**lue – kvanti
- Ezüst – érzékenyebb, de NEM kvanti

Trükkök:

- funkciós csoportra specifikus festés
- differenciál festés

DIGE

Differential
In
Gel
Electrophoresis



2D-kromatográfia/A

1. dimenzió: kation - ioncsere
2. dimenzió: fordított fázisú HPLC (savas pH)

Követés: 215 nm-en → peptidkötés

280 nm-en → Tyr, Trp

2D-kromatográfia/B

1. dimenzió: fordított fázisú HPLC, bázikus pH
2. dimenzió: fordított fázisú HPLC, savas pH

Követés: 215 nm-en → peptidkötés

280 nm-en → Tyr, Trp

2D-kromatográfia/C

Számos variáció – gél-szűrés, kation- vagy anion-csere, hidrofil kölcsönhatáson alapuló kromatográfia, affinitás-kromatográfia stb.

Nyilván semmi hidrofób nem érvényesül ilyen körülmények között...

Reprodukálhatóság degradálódik az oszlop életkorával!

Ki is kell találni mit is válogattunk szét!

- Edman szekvenálás

→ N-terminuson blokkolt fehérjékkel, keverékkel nem megy

- Western-blot

→ tudni kell, hogy mit keresünk, és kell jó ellenanyag

alapötlet

- Egy bizonyos fehérje adott specificitású enzimmel emésztve szekvenciájára jellemző hasítási termékeket fog produkálni

- Ha a valóságos emésztményt az adatbázis „in-silico” eredményeivel összevetjük azonosítani tudjuk a fehérjét

Az összehasonlítás alapja csak valami biztosan megjósolható és egyértelműen meghatározható tulajdonság lehet:

TÖMEG

Q5SY68 **Protein S100-A7-like 2** MW: **11302**

P01842 **Ig lambda chain C regions** MW: **11237**

m/z (mi)
590.2893
703.3621
807.3553
990.5102
1654.738
1743.859
1986.018
2154.13

m/z (mi)	Start	End	Sequence
491.2096	63	66	(K) DDNK(D)
897.3697	19	26	(R) QYSGDDGR(M)
901.4812	1	8	(-) MNIPLGEK(V)
1194.637	9	18	(K) VMLDIVAMFR(Q)
1243.556	89	101	(K) IMHGVAPCSGGSQ(-)
1248.615	27	37	(R) MDMPGLVNLMK(E)
1384.62	38	49	(K) ENFPNFLSGCEK(S)
1385.625	50	61	(K) SDMDYLSNALEK(K)
2201.124	70	88	(K) VNYSEFLSLLGDITIDHHK(I)

A tömegspektrometria előnyei

- alkalmazható keverékekre
- blokkolt fehérjékre is
- kovalens módosítások azonosíthatóak

Detektálási érzékenység

Edman szekvenálás – kb. 1 pmól - 10^{-12} mól

MS – kb. 5-50 fmól - $5x\ 10^{-15}$ - 10^{-14} mól

Western blot – *kevesebb*, mint 10^{-15} mól

(Többnyire) „parciális” információból azonosítjuk a fehérjéket!

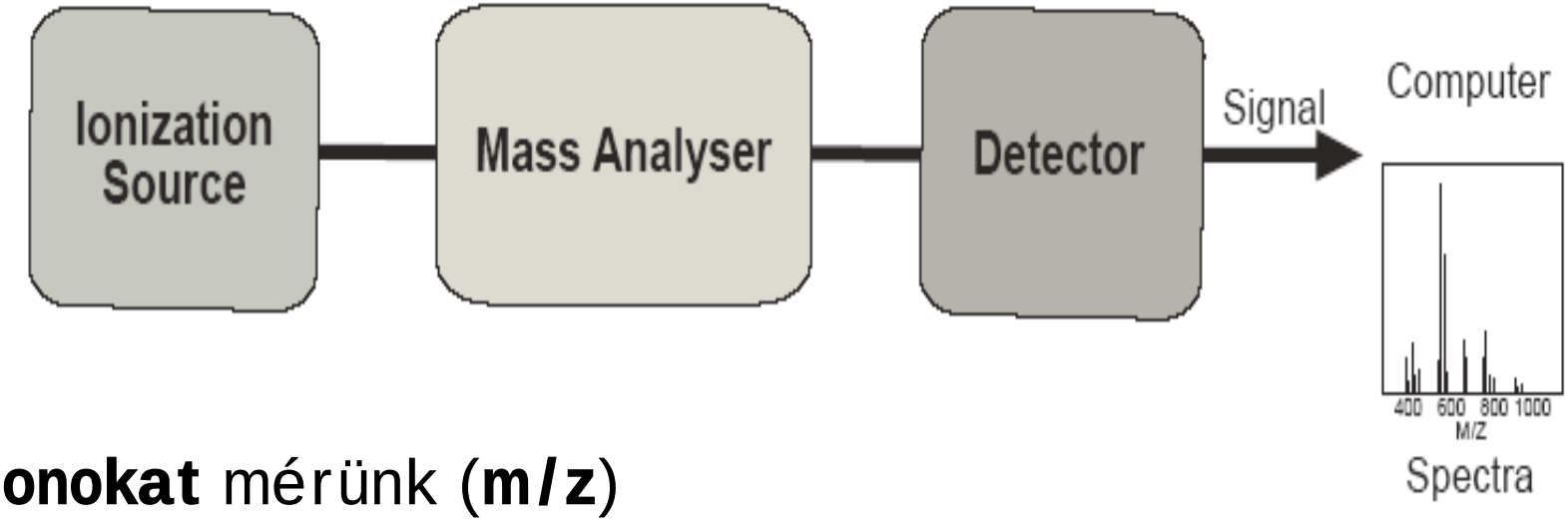
MS-alapú Proteomika

Fordított bioinformatika:

mérési adatok + adatbázisok + algoritmusok → biológiai eredmények

A megbízható kiértékeléshez ismerni kell az analitikai módszert!!!

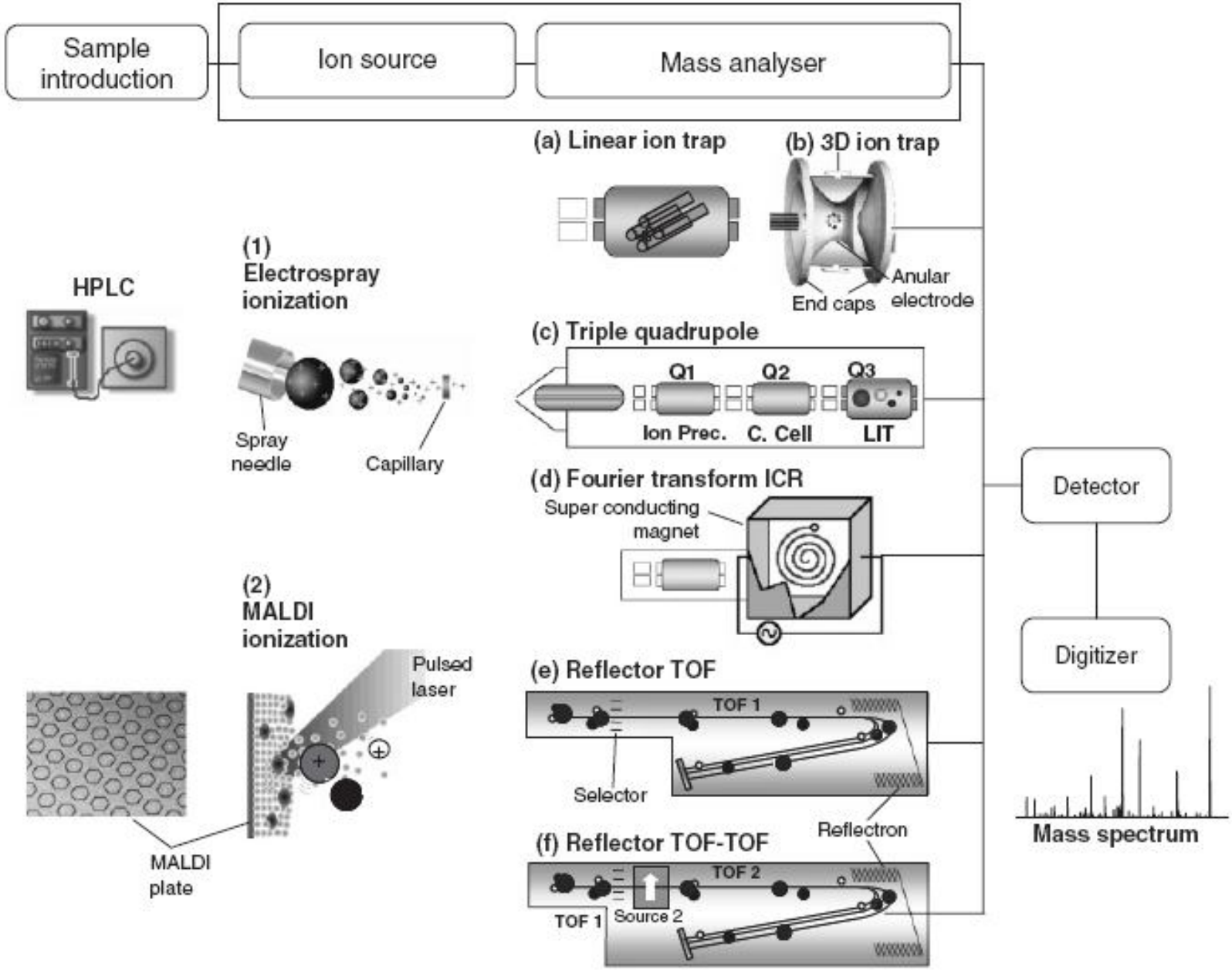
Tömegspektrometria

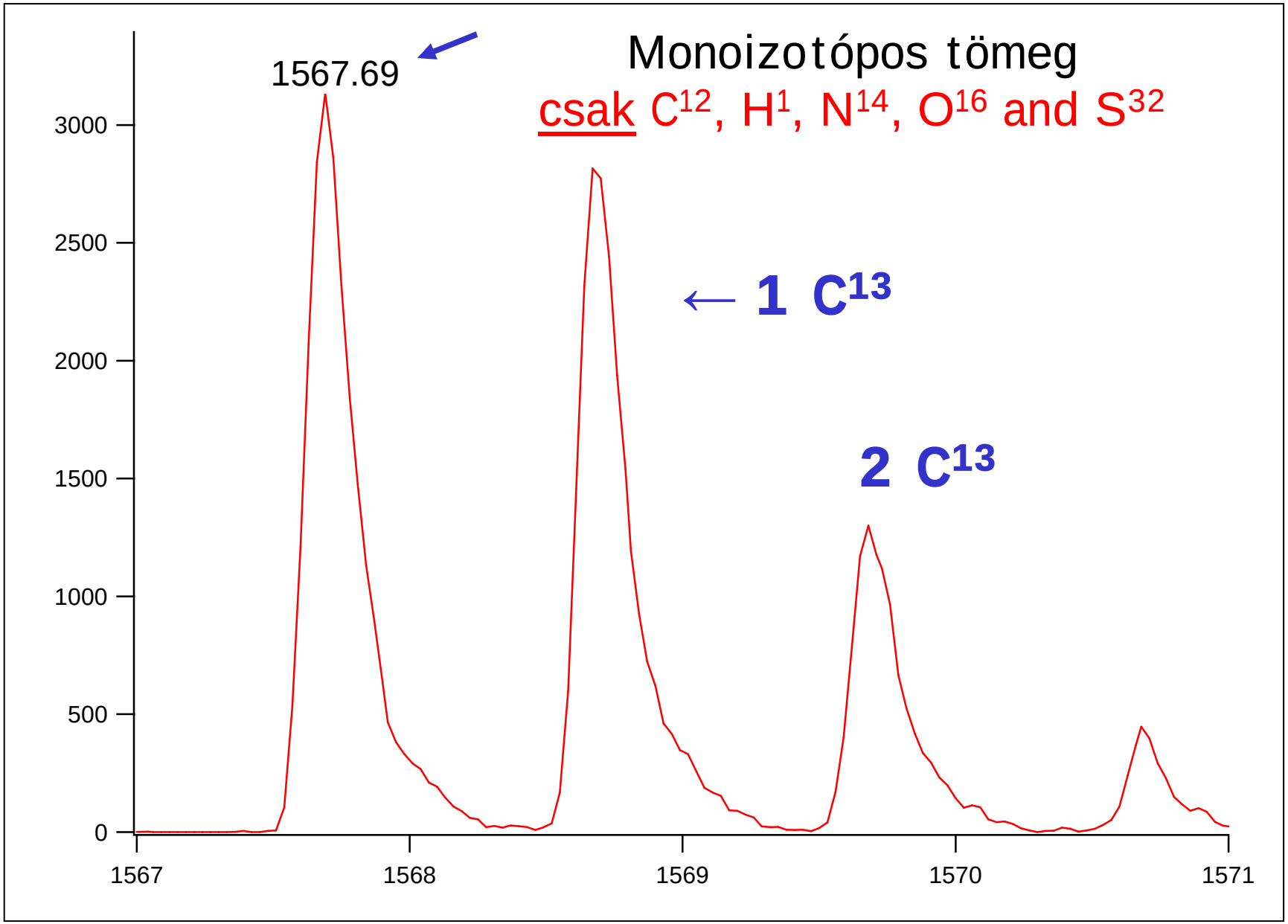


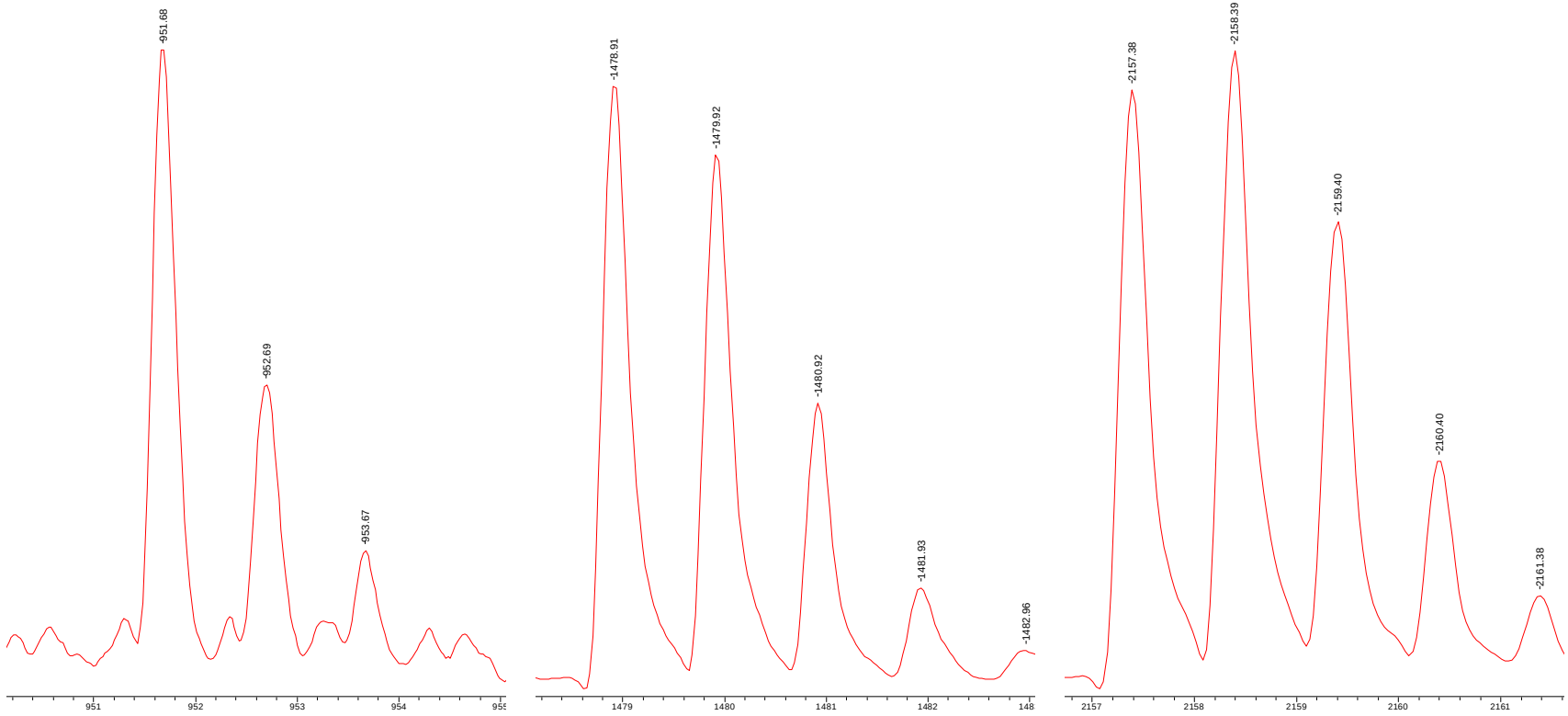
- **Ionokat** mérünk (**m/z**)
  - **M**atrix **A**ssisted **L**aser **D**esorption **I**onization
  - **E**lectro**s**pray **I**onization – többszörösen töltött ionok
- Nagy vákuumban
- **q**uadrupole, **T**ime-**o**f-**F**light, ioncsapda

MS → MH^+ adatok

MS-MS (CID, CAD, HCD, PSD, ETD ...) → szekvencia info


Sample introduction
Ion source
Mass analyser
(a) Linear ion trap
(b) 3D ion trap
Anular electrode
End caps
HPLC
(1) Electrospray ionization
Spray needle
Capillary
(c) Triple quadrupole
Q1
Q2
Q3
Ion Prec.
C. Cell
LIT
(d) Fourier transform ICR
Super conducting magnet
Detector
Digitizer
(2) MALDI ionization
Pulsed laser
MALDI plate
(e) Reflector TOF
TOF 1
Selector
Reflectron
Mass spectrum
(f) Reflector TOF-TOF
TOF 2
TOF 1
Source 2


Monoizotópos tömeg
csak C12, H1, N14, O16 and S32
1567.69
← 1 C13
2 C13
3000
2500
2000
1500
1000
500
0
1567
1568
1569
1570
1571



m/z 952.69 1478.9 2157.38

Tömegmérés pontossága

0.1 Da mérési hiba *egész jó* 2 kDa-nál,

DE *katasztrofális* egy kismolekulánál

Relatív értéket adjunk meg!

part **p**er **m**illion

Milyen pontosan tudunk mérni?

belső standarddal 20 ppm-en belül

± 0.02 Da @ 1000

külső standarddal 200 ppm-en belül

± 0.2 Da @ 1000

intakt fehérjékre 0.1%-on belül

± 10 Da @ 10 000

Ezt tudja az SzBK MALDI készüléke

Milyen pontosan tudunk mérni?

Van olyan masina, ami külső kalibrációval is 2 ppm-en belül mér,

Azaz a hiba <0.002 Da @m/z 1000

Elég ez?

Mennyire kellene pontosan mérni a peptideket ?

MH^+ = 1296.4 ± 0.3 ppm

- 4 különböző elemi összetétel
- 646 különböző aminosav-összetétel
- 12622670280 különböző szekvencia

Ile/Leu ThrVal/SerLeu GlyGlu/AlaAsp

Azért lehet csak tömegekből is azonosítani...

Ez az ún. Peptide Mass Fingerprint – alapú azonosítás

Nyilván akkor, ha a fehérje-frakcionálás jól működik: mondjuk, van egy 2D gélünk

Minta-előkészítés

Legyen a fehérje hozzáférhető a hasításra:

→ denaturálás

→ diszulfid-hidak bontása

Emésztés / kémiai hasítás

→ sótalanítás / tisztítás

Specifikus hasítások

- tripszin - Lys↓ Arg ↓
- endoproteáz Lys-C - Lys ↓
- endoproteáz Glu-C - Glu ↓ (Asp ↓)
- endoproteáz Asp-N - ↓Asp (↓Glu)

pH kb. 7.5 - 8

- CNBr (+75 %-os TFA) - Met ↓

A tömegspektrometria többnyire tripszint használ

- Minimum egy bázikus aminosav/peptid
- És tudjuk, hogy hol van (C-terminus)
- Megbízható
- Olcsó

Megtörtént az emésztés...Jön az analízis!

Mondjuk, MALDI

MS-mérés frakcionálatlan elegyben

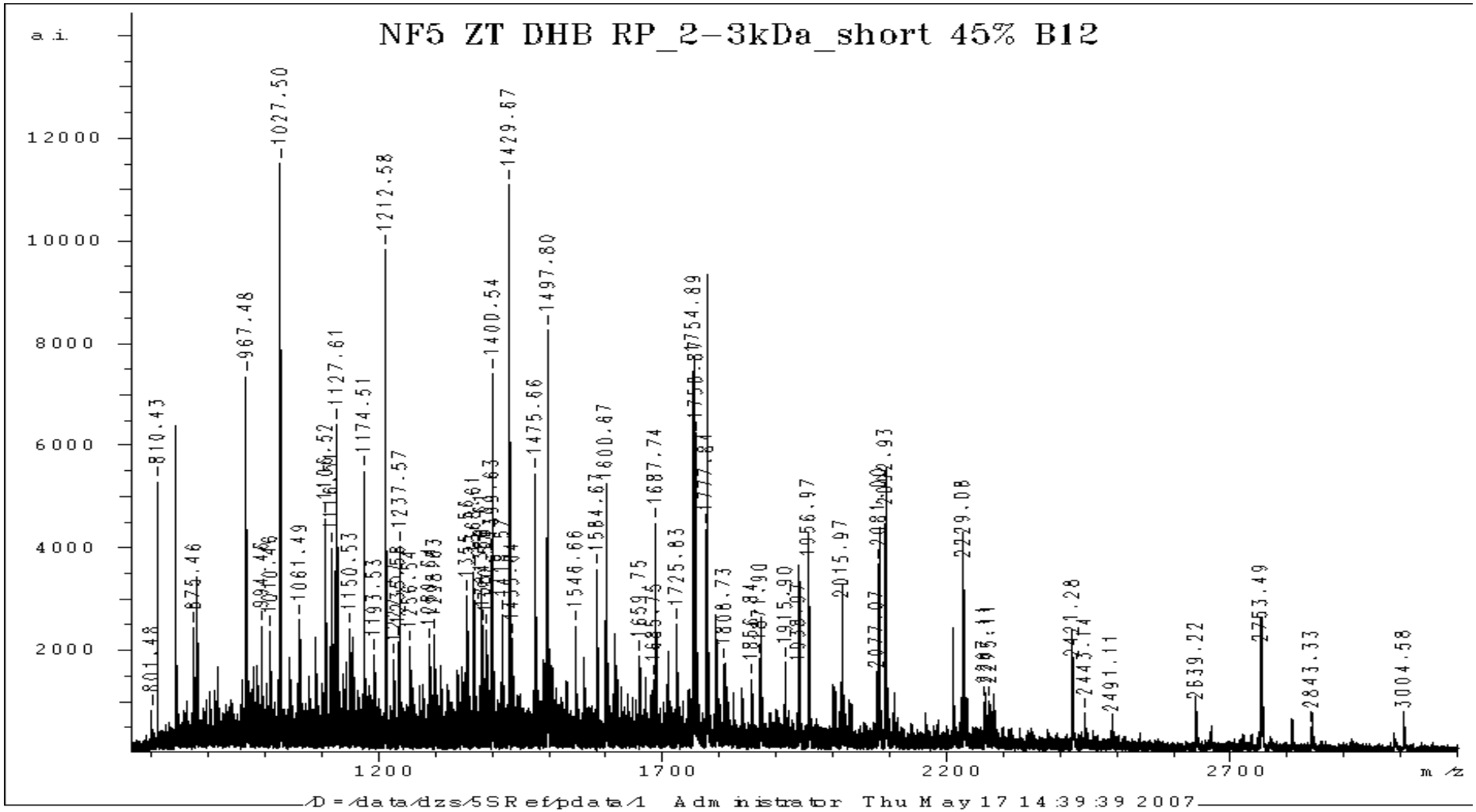
→ egy jellemző (?) MH^+ sorozat

↓ **MS-Fit**

Peptide **M**ass **F**ingerprint – alapú fehérjeazonosítás

Mi lehet a baj?

→ aspecifikus hasítás, kovalens módosítás, szennyezés stb.


NF5 ZT DHB RP_2-3kDa_short 45% B12
a.i.
12000
10000
8000
6000
4000
2000
1200
1700
2200
2700
m/z
801.48
810.43
875.46
967.48
1027.50
1061.49
1127.61
1150.53
1174.51
1193.53
1212.58
1237.57
1400.54
1429.67
1475.66
1497.80
1546.66
1584.67
1600.87
1659.75
1687.74
1725.83
1754.89
1808.73
1915.90
1956.97
2015.97
2229.08
2421.28
2491.11
2639.22
2753.49
2843.33
3004.58
/D:/data/dzs/5SRef/pdata/1 Administrator Thu May 17 14:39:39 2007

PMF lekereső programok

- ProteinProspector MS-Fit
- Mascot
- Aldente
- ProFound

ProteinProspector

v 4.27.1

Proteomics tools for mining sequence databases in conjunction with Mass Spectrometry experiments.

ProteinProspector Asia Pacific

ProteinProspector London

This server uses the IBM AIX version of Protein Prospector

ProteinProspector Tools

Database Search Programs

MS-Fit	MS-Tag	MS-Homology	MS-Bridge
MS-Fit Upload	MS-Seq	MS-Pattern	MS-NonSpecific

Peptide / Protein MS Utility Programs

MS-Digest	MS-Product
MS-Isotope	MS-Comp

Database Management

DB-Stat

Administration/Help

Administering ProteinProspector
User's Manual
FAQ
Bug Listing
ProteinProspector Revision History
ProteinProspector Automation Guidance

Useful Tables

- Mutation Mass Shifts
- Dipeptide Masses
- Trypsin Autolysis Products

Publications
Useful Links

Questions/comments email: ppadmin@cgl.ucsf.edu

MS-Fit (by Peter Baker and Karl Clauser) Instructions

A peptide mass fingerprinting tool from the UCSF Mass Spectrometry Facility that tries to fit a user's mass spectrometry data to a protein sequence in an exis
suggest the identity of the user's protein. The MS input data should be generated by analyzing the peptides produced by the enzymatic digestion of a user's pr

ProteinProspector Home	MS-Tag	MS-Seq	MS-Pattern	MS-Bridge	MS-Fit at UCSF (San Francisco)
MS-Digest	MS-Product	MS-Comp	DB-Stat	MS-Homology	MS-Isotope

Database: NCBInr.10.21.2003 DNA Frame translation: 3
Search Hits: ☐ From: msfit Filename: lastres1
Output Type: HTML Save Hits to file: ☐ Filename: lastres1
Accession Numbers:

Digest: Trypsin
Max. # of missed cleavages: 1
Cys modified by: acrylamide
N term: Hydrogen C term: Free Acid

Species: All
Species Codes:
MW of Protein: (from 1000 Da to 100000 Da) All ☐
Protein pI: (from 3.0 to 10.0) All ☑

Names:

Start Search

Maximum Reported Hits: 5
Sample ID (comment): Magic Bullet digest
Sort By: Score Sort
Min. # peptides required to match: 4
Report MOWSE Scores: ☑ Pfactor: 0.4
Chem Score: ☐ Met Ox Factor: 1.0

Possible Modifications Mode (default)
Peptide N-terminal Gln to pyroGlu
Oxidation of M
Protein N-terminus Acetylated
Acrylamide Modified Cys
User Defined Modification 1:
Phosphorylation of S, T and Y

OR

Homology Mode (select any mode but identity)
Search mode: identity
Min. # matches with NO AA substitutions: 1

Chem Score: ☐ Met Ox Factor: 1.0

Search mode: identity

Min. # matches with NO AA substitutions: 1

Instrument: MALDI-TOF

Peptide masses are: monoisotopic

Mass Tolerance: +/- 50 ppm

Systematic Error: 0

Contaminant Masses:

Data Format: PP M/Z Charge

Display Graphs: ☐

Data Paste Area

```
676.3718
825.4571
1017.5370
1026.5964
1040.6121
1105.4550
1196.7132
1218.5796
1243.5887
1255.6538
1260.5928
1266.6281
1276.5836
1288.6094
1305.6310
1388.6975
```

Other peptide-mass fingerprinting servers:

Peptide Search | Mascot | ProFound | PeptIdent | MultiIdent | MassSearch | MOWSE | PepMapper

Prediction of potential protein post-translational modifications and single amino acid substitutions:

FindMod

Data Set 1 Results

MS-Fit search selects **2284** entries (results displayed for top **5** matches).

Results Summary

	MOWSE Score	#/23(%) Masses Matched	% Cov	% TIC	Mean Err ppm	Data Tol ppm	MS-Digest Index #	Protein MW (Da)/pI	Accession #	Species	Protein Name
1	1.378e+08	17 (73)	61.0	73.9	-3.07	12.0	130345	28432/5.6	245563	BOS=CATTLE, FRIESIAN-HOLSTEIN MALE CALVES AGED 2-4 WEEKS, PEPTIDE, 247 AA	apolipoprotein A
2	1.341e+08	17 (73)	57.0	73.9	-3.07	12.0	1335906	30277/5.7	113988	BOVIN	Apolipoprotein A
3	1.086e+07	15 (65)	53.0	65.2	-2.41	12.2	879310	30272/5.8	27806311 M	UNREADABLE	gi|27806311|ref| apolipoprotein A
4	1226	4 (17)	79.0	17.4	-38.4	15.8	868726	3312/10.4	913734	CAULIFLOWER MOSAIC VIRUS	aphid transmissi
5	1169	4 (17)	76.0	17.4	-38.4	15.8	456837	3475/10.2	913729	CAULIFLOWER MOSAIC VIRUS	aphid transmissi

130345	x	x	x	x	x	.	x	x	.	x	x	x	.	x	x	x	x	.	.	x	x	.	x
1335906	x	x	x	x	x	.	x	x	.	x	x	x	.	x	x	x	x	.	.	x	x	.	x
879310	x	x	x	x	x	.	x	x	.	x	x	x	.	.	.	x	x	.	.	x	x	.	x
868726	.	.	.	.	.	.	.	.	.	x	x	.	x	.	.	x	.	.	.	.	.	.	.
456837	.	.	.	.	.	.	.	.	.	x	x	.	x	.	.	x	.	.	.	.	.	.	.

Detailed Results

1. 17/23 matches (73%).
Acc. #: 245563 **Species:** BOS=CATTLE, FRIESIAN-HOLSTEIN MALE CALVES AGED 2-4 WEEKS, PEPTIDE, 247 AA **Name:** apolipoprotein A
Index: 130345 **MW:** 28432 Da **pI:** 5.6

m/z Submitted	MH^+ Matched	Delta ppm	Modifications	Start	End	Missed Cleavages	Database Sequence
676.3718	676.3630	13		83	88	0	(K)ETASLR (Q)
825.4571	825.4583	-1.4		159	165	0	(R)AHVETLR (Q)
1017.5370	1017.5369	0.079		124	132	0	(K)VAPLGEEFR (E)
1026.5964	1026.5948	1.6		146	154	0	(K)LSPLAQELR (D)
1040.6121	1040.6104	1.6		211	219	0	(K)AKPVLEDLR (Q)
1196.7132	1196.7254	-10		220	230	0	(R)QGLLPVLESLK (V)
1218.5796	1218.5755	3.4		188	199	0	(K)EGGGSLAEYHAK (A)
1255.6538	1255.6574	-2.9		18	28	0	(K)DFATVYVEAIK (D)
1260.5928	1260.6013	-6.8		113	121	0	(K)WHEEVEIYR (Q)
1266.6281	1266.6370	-7.0		102	111	0	(K)VQPYLDEFQK (K)
1288.6094	1288.6174	-6.2	pyroGlu	166	176	0	(R)QQLAPYSDDLR(Q)
1305.6310	1305.6439	-9.9		166	176	0	(R)QQLAPYSDDLR (Q)
1388.6975	1388.6963	0.88		112	121	1	(K)KWHEEVEIYR (Q)
1398.6805	1398.6905	-7.2		33	45	0	(R)DYVAQFEASALGK (Q)
1482.8101	1482.8208	-7.2		16	28	1	(R)VKDFATVYVEAIK (D)
1576.8138	1576.8223	-5.4		51	64	0	(K)LLDNWDTLASTLSK (V)
2188.0850	2188.1039	-8.6		65	82	1	(K)VREQLGPVTQEFWDNLEK (E)

Click link below to search for cysteine linked fragments.
6 unmatched masses:

Click link below to do a non-specific cleavage search.
6 unmatched masses:

Parameters

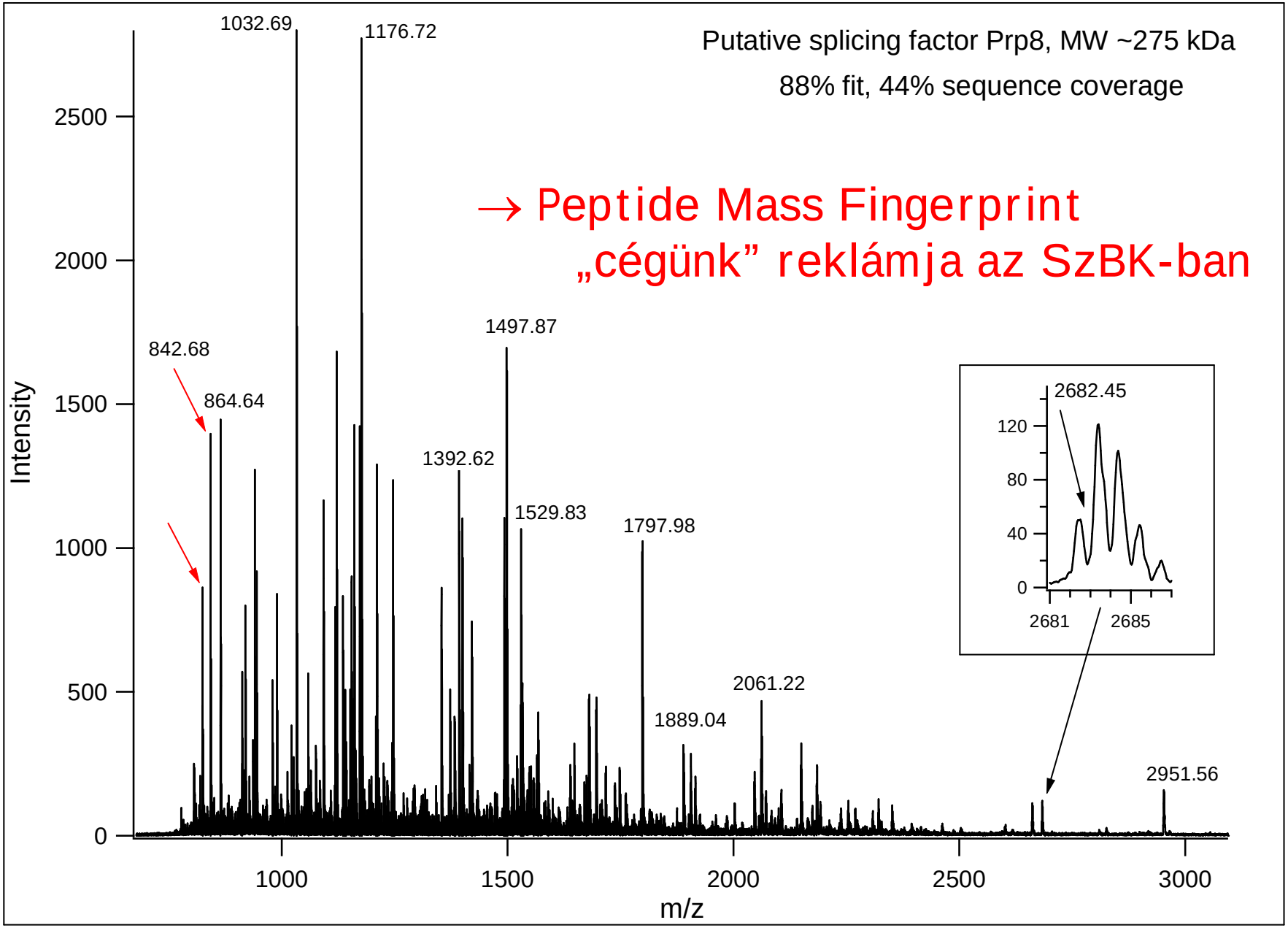
Database: **NCBInr.10.21.2003**
Index Number: **130345**
Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** |
Digest Used: **Trypsin**
Max. # Missed Cleavages: **1**
Peptide N terminus: **Hydrogen**
Peptide C terminus: **Free Acid**
Cysteine Modification: **acrylamide**
Instrument Name: **MALDI-TOF**
Minimum Digest Fragment Mass: **500**
Maximum Digest Fragment Mass: **4000**
Minimum Digest Fragment Length: **5**

Index Number: **130345**
Acc. #: 245563 **Species:** BOS=CATTLE, FRIESIAN-HOLSTEIN MALE CALVES AGED 2-4 WEEKS, PEPTIDE, 247 AA **Name:** apolipoprotein A
pI of Protein: **5.6**
Protein MW: **28432**
Amino Acid Composition: **A24 D17 E27 F6 G10 H5 I4 K22 L34 M1 N4 P8 Q22 R16 S14 T7 V15 W5 Y6**

```
1          11         21         31         41         51         61         71
RHFWQQDDPQ SSWDRVKDFA TVYVEAIKDS GRDYVAQFEA SALGKQLNLK LLDNWDTLAS TLSKVREQLG PVTQEFWDNL
81         91         101        111        121        131        141        151
EKETASLRQE MHKDLEEVKQ KVQPYLDEFQ KKWHEEVEIY RQKVAPLGEE FREGARQKVQ ELQDKLSPLA QELRDRARAH
161        171        181        191        201        211        221        231
VETLRQQLAP YSDDLRQRLT ARLEALKEGG GSLAEYHAKA SEQLKALGEK AKPVLEDLRQ GLLPVLESLK VSILAAIDEA
241
SKKLNAQ
```

The matched peptides cover **61%** (152/247 AA's) of the protein.

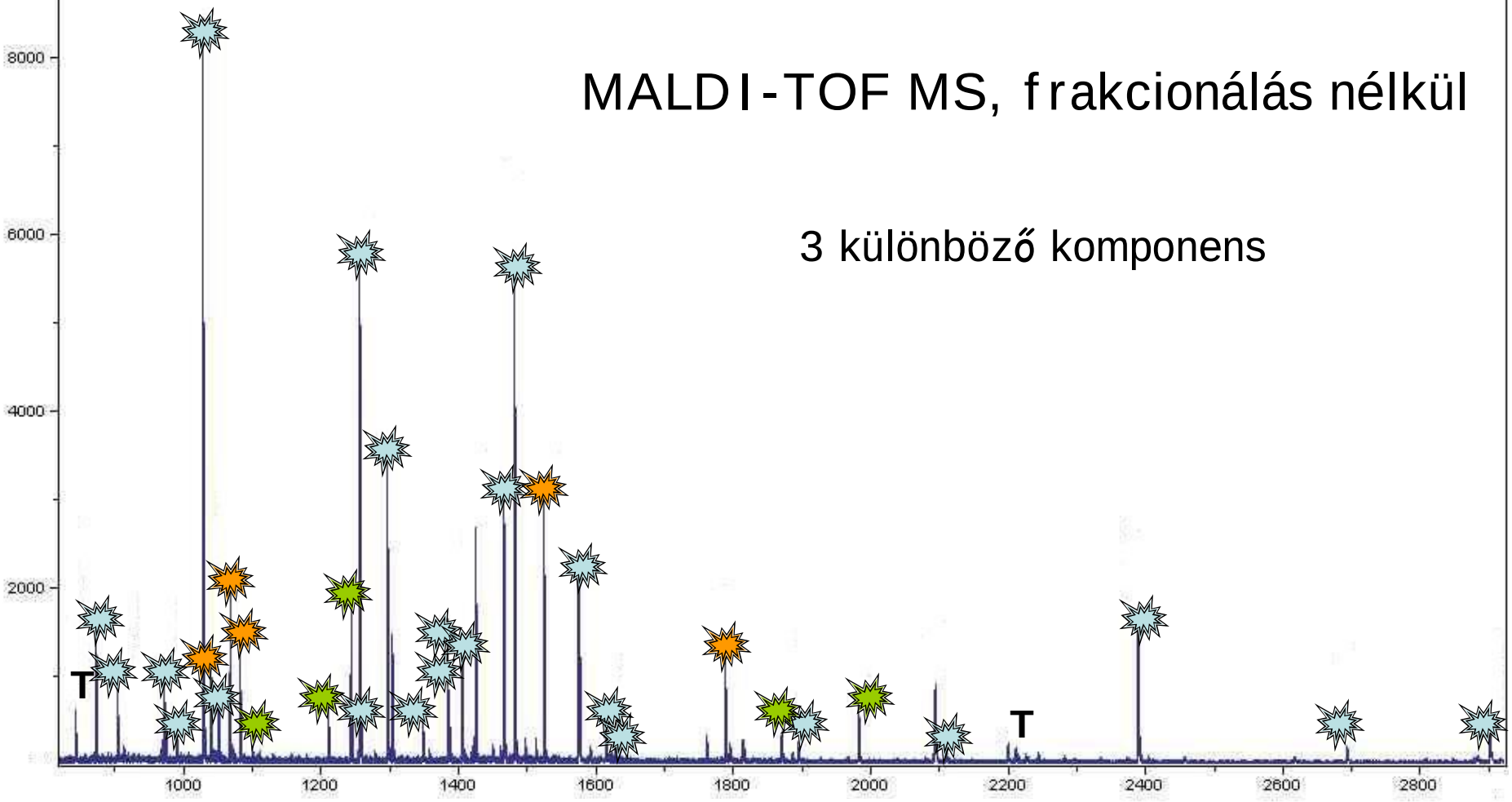
Nagy Ferenc, SzBK



Nem mindig egységes ám a minta!

MALDI-TOF MS, frakcionálás nélkül

3 különböző komponens



Num Unmatched Masses: **26**

Search for disulfide linked peptides.

Do a non-specific cleavage search.

Search for another component.

Azaz szekvenálni muszáj

- a PMF-alapú azonosítást meg kell erősíteni

MIÉRT??? – lásd előbb és később is...

- a „vád tanúja”: az MS-MS kísérlet eredménye

Miért kell még szekvencia info?

- Kovalens módosítások
- Izoformák
- Splice variants
- És ha nincs bent az adatbázisban?

Hogyan tegyünk szert a szükséges információra?

Feladat:

→ meghatározni a komponensek tömegét

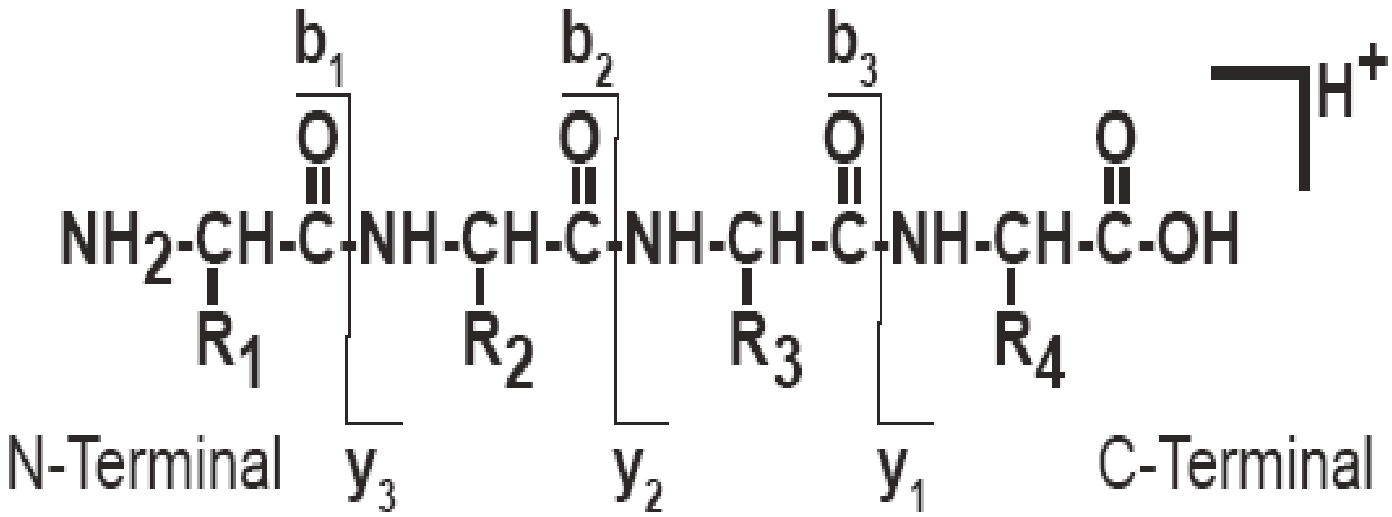
→ egyet fizikailag elkülöníteni

→ „szétverni”

Műszeres megoldás: tandem tömegspektrometria, ionkapu, ioncsapda

Fragmentálás: post-source decay (PSD), ütközéses aktiválás (CID/CAD/HCD), elektron-befogás (ECD/ETD)

Peptid fragmentáció

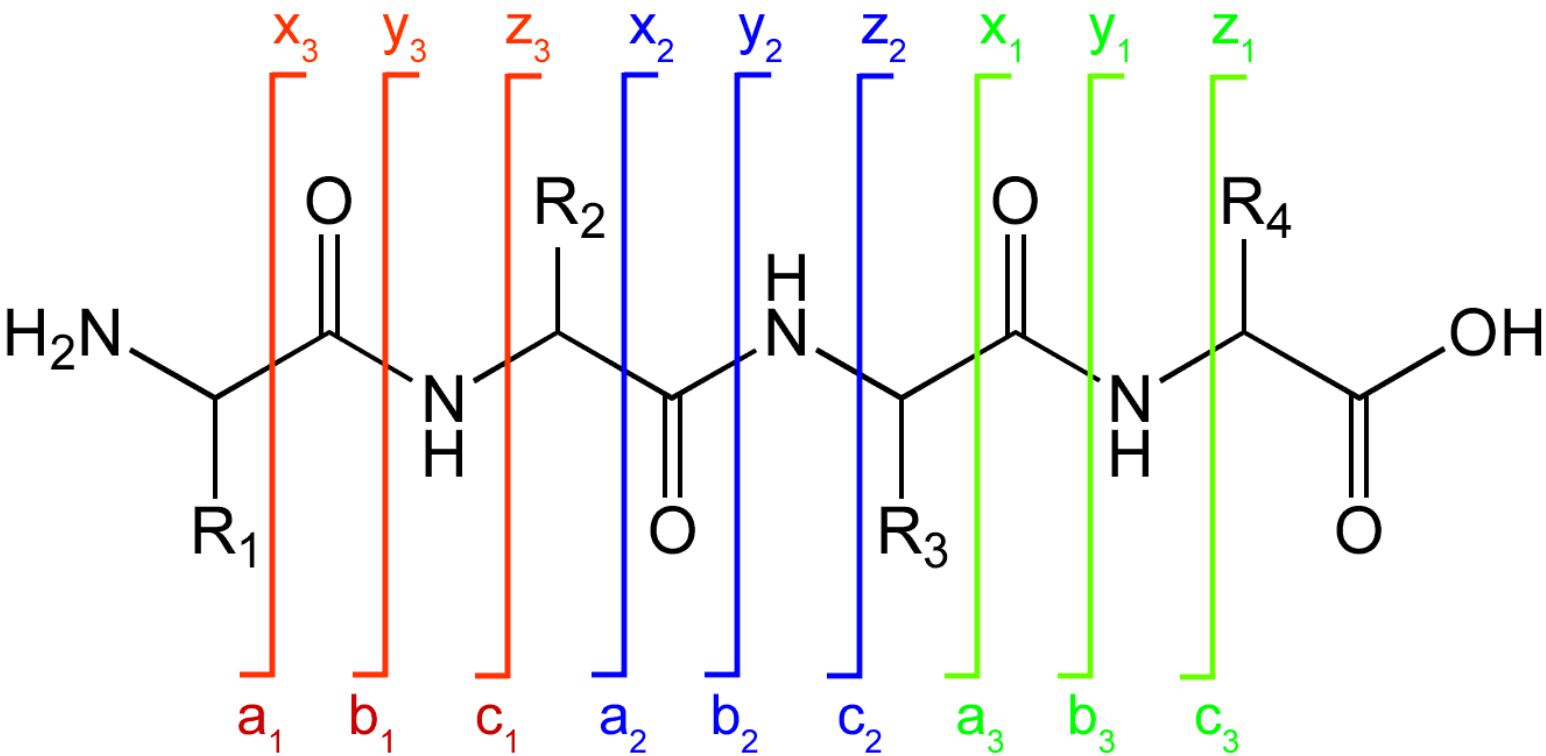


(88) 185 300 463 619 $b_i = \Sigma$ gyöksúly + 1

Ser-Pro-Asp-Tyr-Arg

(637) 550 453 338 175 $y_i = \Sigma$ gyöksúly + 19

A fragmentáció módszer- és készülék-függő!


x3
y3
z3
x2
y2
z2
x1
y1
z1
O
R2
O
R4
H
H2N
N
N
OH
H
R1
O
R3
O
a1
b1
c1
a2
b2
c2
a3
b3
c3

ProteinProspector

v 4.27.1

Proteomics tools for mining sequence databases in conjunction with Mass Spectrometry experiments.

ProteinProspector Asia Pacific

ProteinProspector London

This server uses the IBM AIX version of Protein Prospector

ProteinProspector Tools

Database Search Programs

MS-Fit	MS-Tag	MS-Homology	MS-Bridge
MS-Fit Upload	MS-Seq	MS-Pattern	MS-NonSpecific

Peptide / Protein MS Utility Programs

MS-Digest	MS-Product
MS-Isotope	MS-Comp

Database Management

DB-Stat

Administration/Help

Administering ProteinProspector
User's Manual
FAQ
Bug Listing
ProteinProspector Revision History
ProteinProspector Automation Guidance

Useful Tables

- Mutation Mass Shifts
- Dipeptide Masses
- Trypsin Autolysis Products

Publications
Useful Links

Questions/comments email: ppadmin@cql.ucsf.edu

SHREK - Elméleti ESI-CID spektrum – MS product

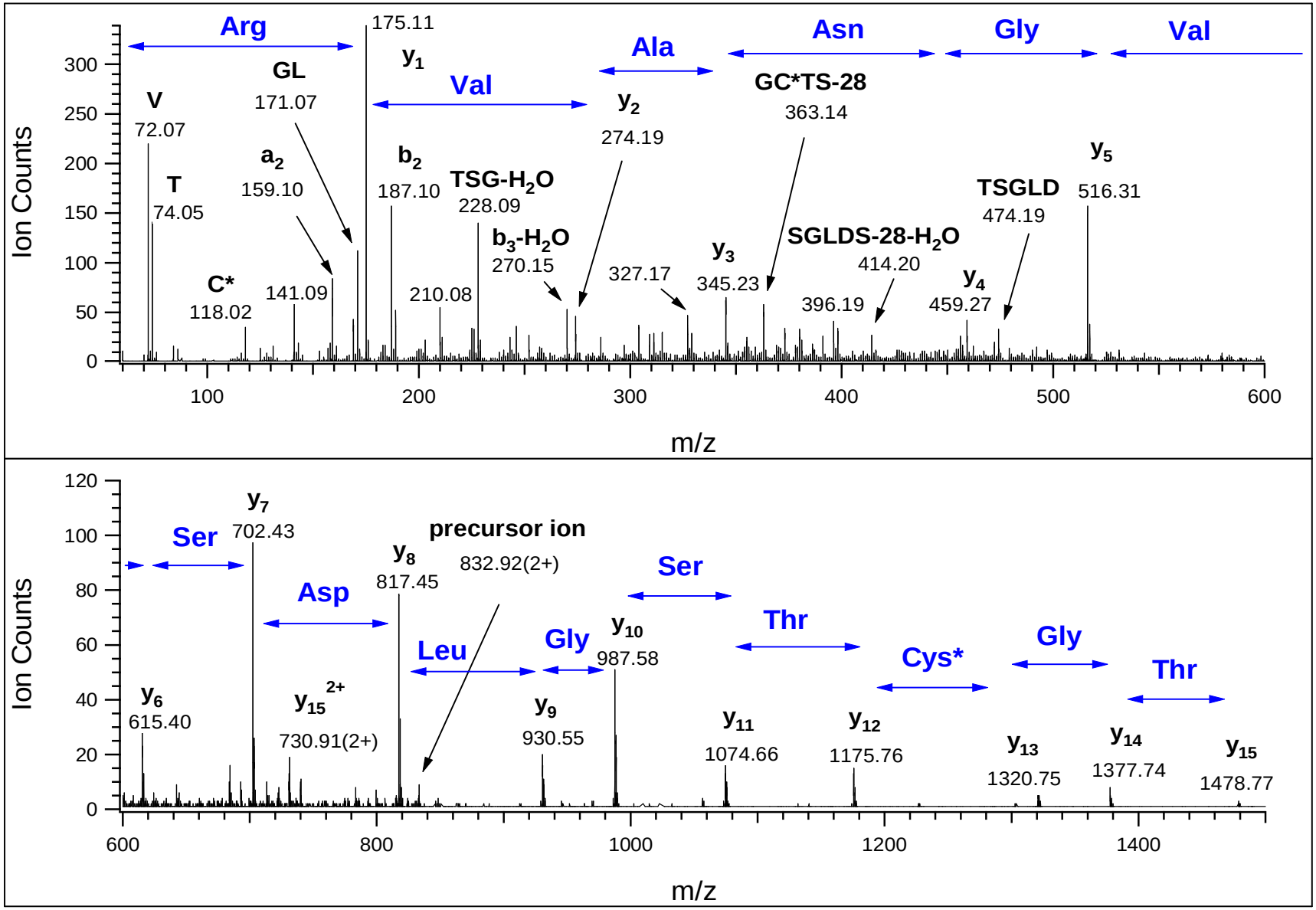
60.0444	**S**	130.0863	**y_1-NH_3**	268.1404	**RE-H_2O**	381.1993	**b_3**	482.2470	**a_4**
70.0651	**R**	138.0662	**H**	269.1244	**RE-NH_3**	395.2150	**HRE-28**	492.2314	**b_4-H_2O**
84.0808	**K**	147.1128	**y_1**	276.1554	**y_2**	399.2099	**b_3+H_2O**	493.2154	**b_4-NH_3**
87.0917	**R**	179.0927	**a_2-H_2O**	277.1408	**HR-NH_3**	405.1993	**HRE-H_2O**	510.2419	**b_4**
100.0869	**R**	197.1033	**a_2**	286.1510	**RE**	406.1833	**HRE-NH_3**	528.2525	**b_4+H_2O**
101.1073	**K**	207.0877	**b_2-H_2O**	294.1673	**HR**	414.2459	**y_3-H_2O**	551.3049	**y_4-H_2O**
102.0550	**E**	225.0982	**b_2**	335.1938	**a_3-H_2O**	415.2300	**y_3-NH_3**	552.2889	**y_4-NH_3**
110.0713	**H**	258.1448	**y_2-H_2O**	336.1779	**a_3-NH_3**	423.2099	**HRE**	569.3154	**y_4**
112.0869	**R**	258.1561	**RE-28**	353.2044	**a_3**	432.2565	**y_3**	638.3369	**MH-H_2O**
126.0913	**K**	259.1288	**y_2-NH_3**	363.1888	**b_3-H_2O**	464.2364	**a_4-H_2O**	639.3209	**MH-NH_3**
129.1022	**K**	266.1724	**HR-28**	364.1728	**b_3-NH_3**	465.2205	**a_4-NH_3**	656.3474	**MH**

SHREK – elméleti ECD fragmensek

131.0946	$\mathbf{z_1}$	242.1253	$\mathbf{c_2}$	353.2050	$\mathbf{a_3}$	416.2383	$\mathbf{z_3}$	527.2690	$\mathbf{c_4}$
197.1039	$\mathbf{a_2}$	260.1372	$\mathbf{z_2}$	398.2264	$\mathbf{c_3}$	482.2476	$\mathbf{a_4}$	553.2972	$\mathbf{z_4}$

Elméleti „spektrum”:
→ lehetséges fragmensek, egyenlő intenzitással

És egy valóságos spektrum



(186)TGCTSGI/LDSVGNAVR

Mihez kezdünk a valóságos MS/MS adatokkal?

MS-Tag Simple (by Karl Clauser and Peter Baker) Instructions

A stripped down MS-Tag input form from the UCSF Mass Spectrometry Facility for people who barely know the difference between a b-ion and a y-ion (and are not afraid to admit it). If you don
monoisotopic and average masses, please consult your local mass spectrometrist before using this program. Note: This form uses exactly the same MS-Tag program as the other HTML forms, it's ju
parameters are not available.

ProteinProspector Home	MS-Fit	MS-Pattern	MS-Seq	MS-Bridge	MS-Tag at UCSF (San Francisco)
MS-Digest	MS-Product	MS-Comp	DB-Stat	MS-Homology	MS-Isotope

MS-Tag

Database: NCBInr.10.21.2003 DNA Frame translation: 3
Search Hits: ☐ From: msfit Filename: lastres1

Digest: Trypsin
Max. # of missed cleavages: 1
Cys modified by: acrylamide

Species: All
Protein MW: (from 1000 to 100000 Da) All ☑

Start Search

Max. % Unmatched Ions 10

Parent tolerance: +/- 100 ppm
Fragment tolerance: +/- 1500

Maximum Reported Hits: 25

Display Graphs: ☐
Fragment-ion Tag
Masses are: monoisotopic

Mass (m/z)	Charge (z)
1040.6121	
70.1	
72.1	
84.1	
86.1	
87.1	
102.1	
112.1	
169.14	
175.14	
197.08	
200.06	

Allowed Fragment-Ion types (specific to instrument type selected)
Instrument: MALDI-TOF

AA Composition (based on immonium and related ions) is calculated automatically

Sample ID (comment): Apo A-1 1040 AKPVLEDLR

MS-MS eredményekkel is lehet keresgélni!

18 Ions used in search: **70.10, 72.10, 84.10, 86.10, 87.10, 102.10, 112.10, 169.14, 175.14, 197.08, 200.06, 280.42, 288.21, 310.45, 358.01, 403.01, 554.11, 883.71 (+/- 1500.00 ppm)**
Max # Unmatched Ions = **1**

Result Summary

Rank	# Unmatched Ions	Measured Sequence	MH^+ Calculated (Da)	MH^+ Error (ppm)	MS-Digest Index #	Protein MW (Da)/pI	Accession #	Species	Protein Name
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	130345	28432/5.6	245563	BOS=CATTLE, FRIESIAN-HOLSTEIN MALE CALVES AGED 2-4 WEEKS, PEPTIDE, 247 AA	apolipoprotein A
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	879310	30272/5.8	27806311 M	UNREADABLE	gi|27806311|ref|h A1 [Bos taurus]
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	1335906	30277/5.7	113988	BOVIN	Apolipoprotein

Detailed Results

Rank	# Unmatched Ions	Measured Sequence	MH^+ Calculated (Da)	MH^+ Error (ppm)	MS-Digest Index #	Protein MW (Da)/pI	Accession #	Species	Protein Name
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	130345	28432/5.6	245563	BOS=CATTLE, FRIESIAN-HOLSTEIN MALE CALVES AGED 2-4 WEEKS, PEPTIDE, 247 AA	apolipoprotein A
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	879310	30272/5.8	27806311 M	UNREADABLE	gi|27806311|ref|h A1 [Bos taurus]
1	0	(K)AKPVLEDLR (Q)	1040.6104	1.6	1335906	30277/5.7	113988	BOVIN	Apolipoprotein

Fragment Ion (m/z)	70.10	72.10	84.10	86.10	87.10	102.10	112.10	169.14	175.14	197.08	200.06	280.42
Ion Type Delta ppm	P 490 R 490	V 259	K 222	L 35	R 89	E 436	R 112	PV-28 35	y_1 117	PV -249	b_2 -399	b_3-NH_3 906

Fragment Ion (m/z)	288.21	310.45	358.01	403.01	554.11	883.71
Ion Type Delta ppm	y_2 22	PVL 764	EDL -423 LED -423	y_3 -547	PVLED -311	b_8+H_2O -904

Független PMF és MS-MS azonosítás az igazi!

A proteomika „Achilles sarka”?

Elméleti azaz megjósolt „adatokon” alapuló azonosítás

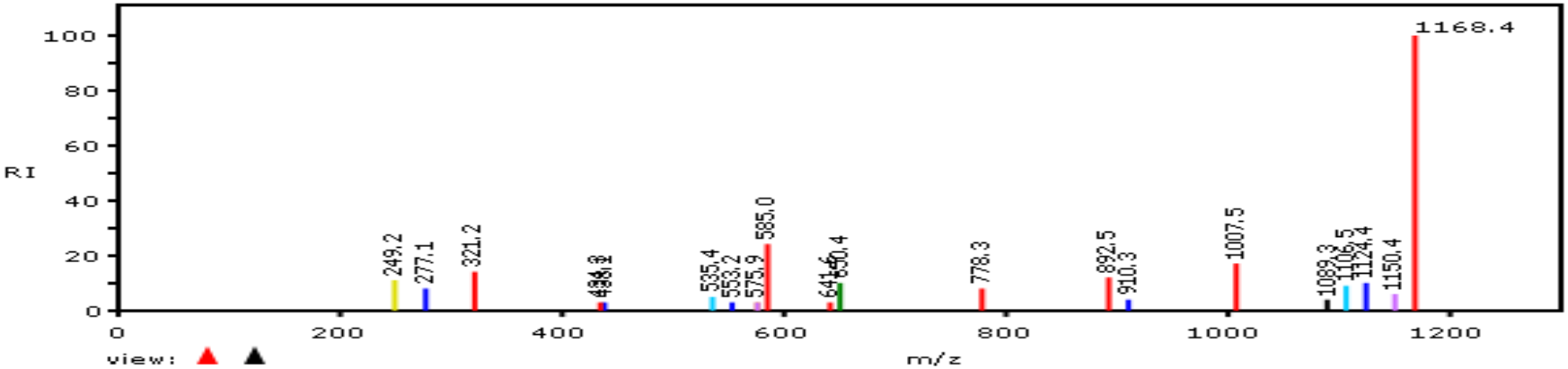
Lehet spektrum-könyvtárhoz hasonlítani is

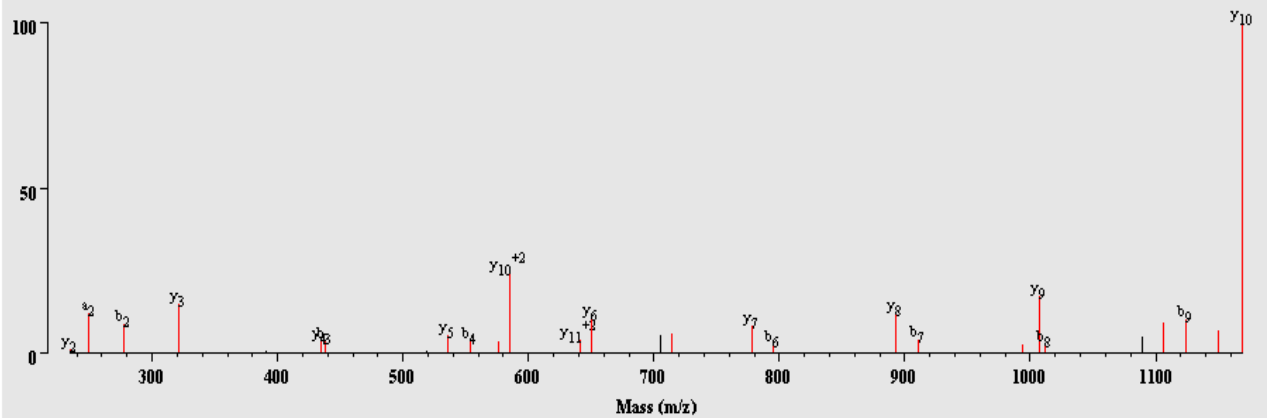
- De csak ugyanolyan készülékkel felvett adatokkal működik jól

Example of Spectrum not Matched in Library Search

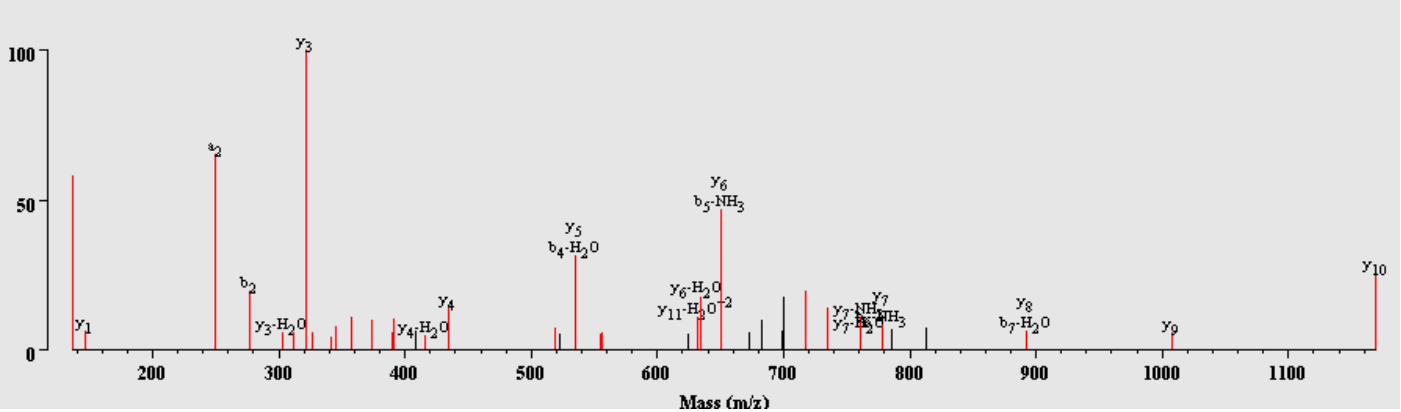
gpmDB current contains **17158** entries for *YICDNQDTISSK*
Your result is compared against the best spectra available.
cos(θ) = 1.0 is the best possible match,
cos(θ) = 0.0 is the worst possible match.

Y I C D N Q D T I S S K





LTQ



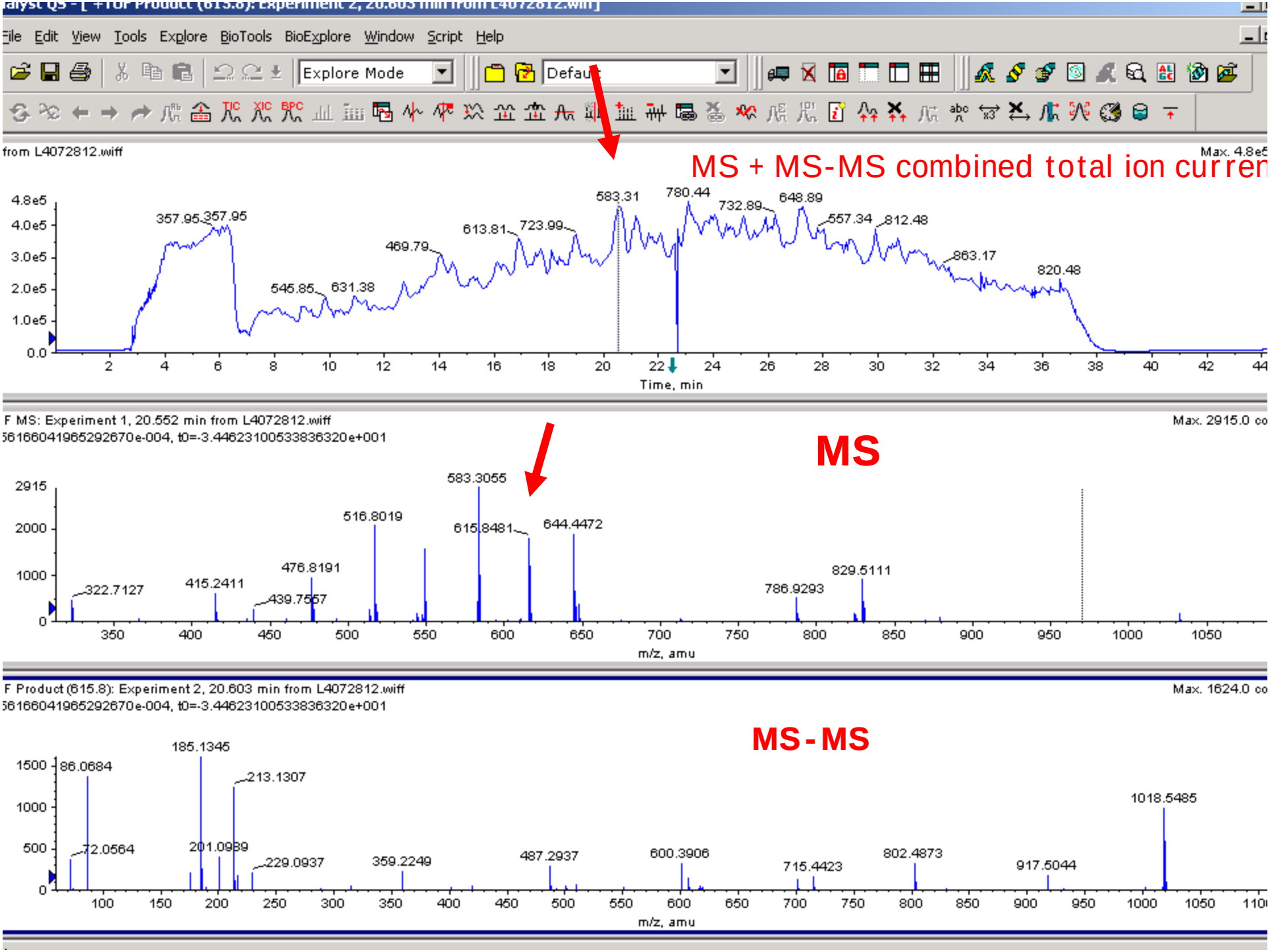
QSTAR

Elegánsabb az on-line LC-MS/MS

A nanoHPLC oszlopról minden megy a tömegspektrométerbe (250 nL/min)

ún. „adatfüggő” adatgyűjtés

1) MS mérés
2) Komputer kiválasztja a legintenzívebb iont
3) MS-MS mérés


File Edit View Tools Explore BioTools BioExplore Window Script Help
Explore Mode
Default
from L4072812.wiff
Max. 4.8e5
MS + MS-MS combined total ion curren
357.95
357.95
545.85
631.38
469.79
613.81
723.99
583.31
780.44
732.89
648.89
557.34
812.48
863.17
820.48
Time, min
F MS: Experiment 1, 20.552 min from L4072812.wiff
Max. 2915.0 co
MS
583.3055
516.8019
615.8481
644.4472
476.8191
322.7127
415.2411
439.7557
786.9293
829.5111
m/z, amu
F Product (615.8): Experiment 2, 20.603 min from L4072812.wiff
Max. 1624.0 co
MS-MS
185.1345
86.0684
213.1307
1018.5485
72.0564
201.0989
229.0937
359.2249
487.2937
600.3906
715.4423
802.4873
917.5044
m/z, amu

Lekeresés csak MS-MS adatokkal

„script” – az *összes* MS-MS adatok megfelelő formába rendezve

az egyes spektrumokkal független lekeresés

az eredményeket viszont fehérjénként összegezzük

LC/MS/MS lekereső programok

- ProteinProspector Batch-Tag
- Mascot
- Sequest
- OMSSA
- SpectrumMill
- .....és még sokan mások

Database Search Parameters

- Enzyme specificity considered
- # of missed cleavages permitted
- Fixed modification(s) (residue specificity)
- Variable modifications (residue specificity)
- Mass tolerance for precursor ions
- Mass tolerance for fragment ions

Search engines work by:

1. Determining a list of potential peptides from a database that can be formed by the specified parameters and have the correct precursor mass.
2. Determining scores for the matches of the fragment peaks to each of these peptides.

- Changing search parameters will change the number of the potential hits.
- For software using probabilities or Expectation values (E-values) for scoring, it will change the scores.

Mascot Search Results

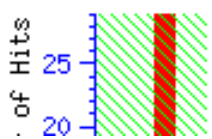
User : kati
Email : folkl@cgl.ucsf.edu
Search title : d:\PE Sciex Data\Projects\MALDI\Data\L4012605.wiff : KFM#2 mouse small
MS data file : C:\Temp\mas156.tmp
Database : NCBInr 20040120 (1607648 sequences; 528681616 residues)
Timestamp : 26 Jan 2004 at 21:12:06 GMT

Significant hits:

gi|2136983	Ig gamma heavy chain constant region - rabbit (fragment)
gi|457366	Ig gamma H-chain C-region [Oryctolagus cuniculus]
gi|12805321	Sec13r protein [Mus musculus]
gi|418694	serum albumin precursor [validated] - bovine
gi|1351908	Serum albumin precursor (Allergen Fel d 2)
gi|3319897	albumin [Canis familiaris]
gi|51653	unnamed protein product [Mus musculus]
gi|136429	Trypsin precursor
gi|16754822	immunoglobulin heavy chain constant region [Sylvilagus cunicularis]
gi|191765	alpha-fetoprotein
gi|125138	Ig kappa-B4 chain C region
gi|113582	Serum albumin precursor
gi|67548	trypsin (EC 3.4.21.4) II precursor - rat
gi|2358072	trypsinogen 7 [Mus musculus]
gi|320525	hypothetical protein (Rv2744c region) - Mycobacterium tuberculosis
gi|13541048	Glycosidase [Thermoplasma volcanium]
gi|29135112	ORF176 [Pseudomonas phage phiKZ]

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 43 indicate identity or extensive homology (p<0.05).



1. gi|2136983 **Mass:** 20734 **Score:** 292 **Peptides matched:** 9

Ig gamma heavy chain constant region - rabbit (fragment)

☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	13	384.73	767.44	767.42	0.02	0	26	3.1	1	GQPLEPK
☑	17	405.77	809.52	809.49	0.03	0	29	1	1	TARPPLR
☑	20	419.77	837.53	837.50	0.03	0	31	0.68	1	ALPAPIEK
☑	37	460.76	919.50	919.46	0.04	0	24	11	1	VYTMGPPR
☑	64	601.84	1201.66	1201.61	0.05	0	64	0.0009	1	LSVPTSEWQR
☑	75	541.29	1620.84	1620.79	0.05	1	19	23	1	VYTMGPPREELSSR
☑	76	545.67	1633.98	1633.89	0.08	0	(43)	0.047	1	VVSTLPIAHQDWLR
☑	77	818.00	1633.99	1633.89	0.10	0	46	0.021	1	VVSTLPIAHQDWLR
☑	81	932.51	1863.01	1862.89	0.11	0	53	0.0087	1	TTPAVLDSDGSYFLYSK

---

2. gi|457366 **Mass:** 35840 **Score:** 288 **Peptides matched:** 9

Ig gamma H-chain C-region [Oryctolagus cuniculus]

☐ Check to include this hit in error tolerant search

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑	10	353.71	705.40	705.38	0.02	0	28	3.5	1	TFPSVR
	13	384.73	767.44	767.42	0.02	0	26	3.1	1	GQPLEPK
	17	405.77	809.52	809.49	0.03	0	29	1	1	TARPPLR
	20	419.77	837.53	837.50	0.03	0	31	0.68	1	ALPAPIEK
☑	23	426.23	850.45	850.42	0.03	0	14	95	1	DTLMISR + Oxidation (M)
	37	460.76	919.50	919.46	0.04	0	24	11	1	VYTMGPPR
	64	601.84	1201.66	1201.61	0.05	0	64	0.0009	1	LSVPTSEWQR
	75	541.29	1620.84	1620.79	0.05	1	19	23	1	VYTMGPPREELSSR
	81	932.51	1863.01	1862.89	0.11	0	53	0.0087	1	TTPAVLDSDGSYFLYSK

☑	17	405.77	809.52	809.49	0.03	0	29	1	1	TARPPLR
☑	20	419.77	837.53	837.50	0.03	0	31	0.68	1	ALPAPIEK
☑	37	460.76	919.50	919.46	0.04	0	24	11	1	VYTMGPPR
☑	64	601.84	1201.66	1201.61	0.05	0	64	0.0009	1	LSVPTSEWQR
☑	75	541.29	1620.84	1620.79	0.05	1	19	23	1	VYTMGPPREELSSR
☑	76	545.67	1633.98	1633.89	0.08	0	(43)	0.047	1	VVSTLPIAHQDWLR
☑	77	818.00	1633.99	1633.89	0.10	0	46	0.021	1	VVSTLPIAHQDWLR
☑	81	932.51	1863.01	1862.89	0.11	0	53	0.0087	1	TTPAVLDSDGSYFLYSK

2. gi|4
Ig g
☐ Chec

Query
☑ 10
13
17
20
☑ 23
37
64
75
81

Top scoring peptide matches to query 81
(Elution from: 21.41 to 21.41 period: 0 experiment: 1 cycles: 1)
Score greater than 28 indicates homology
Score greater than 45 indicates identity
Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
53.0	0.11	1+	gi|2136983	TTPAVLDSDGSYFLYSK
12.5	-0.04			LAPIGVFFLILPNMFR
11.0	-0.09			IGAGAKIGAGSVVLIDVPAR
9.5	0.05			QLKVMDDFIKPYIDR
8.5	0.09			QLAEVFDRSTIDYAPR
8.2	0.02			CIHIIVECANVLVAPR
7.1	0.04			QYGFVPVQMTNLVPVR
6.9	0.16			MANLSYNSNKASSDFSK
5.8	0.06			NKQVDSLVGVNTNNFSK
5.7	0.19			IAIHNATEADDGAYSCR

lation (M)
R
LYSK

Nem igazi adatokhoz, hanem az elméleti spektrumhoz hasonlítunk!

3. gi|12805321 **Mass:** 36014 **Score:** 246 **Peptides matched:** 5
Sec13r protein [Mus musculus]
☐ Check to include this hit in error tolerant search

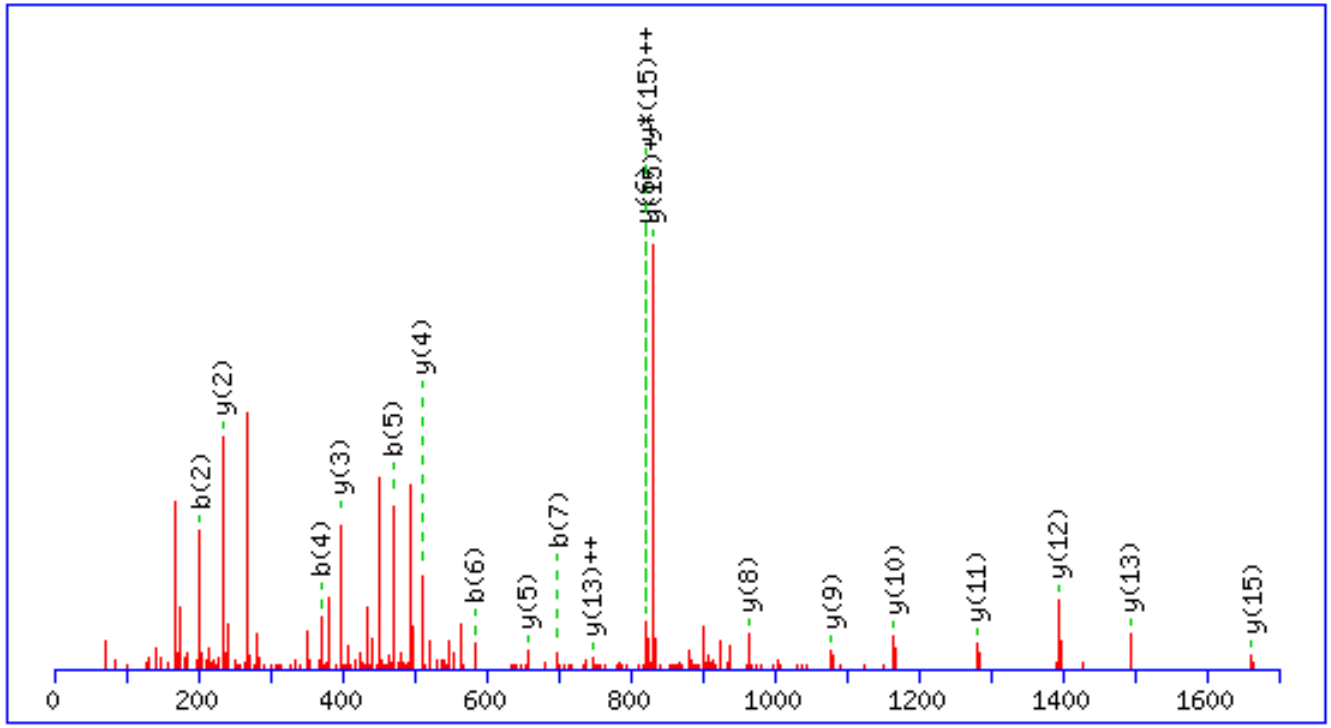
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 5	325.19	648.37	648.36	0.01	0	45	0.12	1	IFDVR
☑ 35	455.23	908.44	908.40	0.04	0	34	0.88	1	LATCSSDR
☑ 57	556.80	1111.59	1111.54	0.05	0	36	0.51	1	LEAHSDWVR
☑ 59	585.36	1168.70	1168.66	0.04	0	49	0.024	1	NGGQILIADLR

Peptide View

MS/MS Fragmentation of **TTPAVLDSDGSYFLYSK**
Found in **gi|2136983**, Ig gamma heavy chain constant region - rabbit (fragment)

Match to Query 81: 1863.007588 from(932.511070,2+)
Elution from: 21.41 to 21.41 period: 0 experiment: 1 cycles: 1
From data file C:\Temp\mas156.tmp

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 0 to 1700 Da



Monoisotopic mass of neutral peptide Mr(calc): 1862.89
Fixed modifications: Carbamidomethyl (C)
Ions Score: 53 **Expect:** 0.0087

Monoisotopic mass of neutral peptide Mr(calc): 1862.89
Fixed modifications: Carbamidomethyl (C)
Ions Score: 53 **Expect:** 0.0087
Matches (Bold Red): 20/128 fragment ions using 61 most intense peaks

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y*	y^{*++}	#
1	74.06	37.53	102.05	51.53	T					17
2	175.11	88.06	**203.10**	102.05	T	1762.85	881.93	1745.83	873.42	16
3	272.16	136.58	300.16	150.58	P	**1661.81**	**831.41**	1644.78	**822.89**	15
4	343.20	172.10	**371.19**	186.10	A	1564.75	782.88	1547.73	774.37	14
5	442.27	221.64	**470.26**	235.63	V	**1493.72**	**747.36**	1476.69	738.85	13
6	555.35	278.18	**583.34**	292.18	L	**1394.65**	697.83	1377.62	689.31	12
7	670.38	335.69	**698.37**	349.69	D	**1281.56**	641.29	1264.54	632.77	11
8	757.41	379.21	785.40	393.21	S	**1166.54**	583.77	1149.51	575.26	10
9	872.44	436.72	900.43	450.72	D	**1079.50**	540.26	1062.48	531.74	9
10	929.46	465.23	957.45	479.23	G	**964.48**	482.74	947.45	474.23	8
11	1016.49	508.75	1044.48	522.75	S	907.46	454.23	890.43	445.72	7
12	1179.55	590.28	1207.55	604.28	Y	**820.42**	410.72	803.40	402.20	6
13	1326.62	663.81	1354.62	677.81	F	**657.36**	329.18	640.33	320.67	5
14	1439.71	720.36	1467.70	734.35	L	**510.29**	255.65	493.27	247.14	4
15	1602.77	801.89	1630.76	815.89	Y	**397.21**	199.11	380.18	190.59	3
16	1689.80	845.40	1717.80	859.40	S	**234.14**	117.58	217.12	109.06	2
17					K	147.11	74.06	130.09	65.55	1

Ig gamma heavy chain constant region - rabbit (fragment)
Found in search of C:\Temp\mas156.tmp

Nominal mass (M_r): **20734**; Calculated pI value: **7.94**
NCBI BLAST search of gi|2136983 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Oryctolagus cuniculus
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|165110 from Oryctolagus cuniculus

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **42%**

Matched peptides shown in **Bold Red**

```
  1 EDDPEVQFTW YINNEQVRTA RPPLREQQFN STIRVVSTLP IAHQDWLRGK
 51 EFKCKVHNKA LPAPIEKTIS KARGQPLEPK VYTMGPPREE LSSRSVSLTC
101 MINGFYPSDI SVEWEKNGKA EDNYKTTPAV LDSDGSYFLY SKLSVPTSEW
151 QRGDVFTCSV MHEALHNHYT QKSISRSPGK
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence	
19 - 25	405.77	809.52	809.49	0.03	0	TARPPLR	(Ions score 29)
35 - 48	545.67	1633.98	1633.89	0.08	0	VVSTLPIAHQDWLR	(Ions score 43)
35 - 48	818.00	1633.99	1633.89	0.10	0	VVSTLPIAHQDWLR	(Ions score 46)
60 - 67	419.77	837.53	837.50	0.03	0	ALPAPIEK	(Ions score 31)
74 - 80	384.73	767.44	767.42	0.02	0	GQPLEPK	(Ions score 26)
81 - 88	460.76	919.50	919.46	0.04	0	VYTMGPPR	(Ions score 24)
81 - 94	541.29	1620.84	1620.79	0.05	1	VYTMGPPREELSSR	(Ions score 19)
126 - 142	932.51	1863.01	1862.89	0.11	0	TTPAVLDSDGSYFLYSK	(Ions score 53)
143 - 152	601.84	1201.66	1201.61	0.05	0	LSVPTSEWQR	(Ions score 64)

Vannak ám „közös” peptidek

- Homológia fajok között
- Fehérje-családok
- Izoformák
- Véletlen egyezés

Sample	Gene	Acc.No.	Protein	Protein MW	Unique Peptides	Coverage
1	At3g02090	Q42290	mitochondrial-processing peptidase beta subunit, (MPP beta)	59	32	55.00%
	At3g16480	O04308	mitochondrial-processing peptidase alpha subunit 2, (MPP alpha-2)	54	1[a]	3.00%
	At2g07727	P42792	Cytochrome b (MTCYB) (COB) (CYTB)	44.5	1[b]	3.30%
2	At1g51980	Q9ZU25	mitochondrial-processing peptidase alpha subunit 1 (MPP Alpha-1)	54[c]	24	40.40%
	At3g16480	O04308	mitochondrial-processing peptidase alpha subunit 2, (MPP alpha-2)	54[c]	1[c](9)	16.40%
3	At5g40810	Q0WNJ4	Cytochrome c1 (CYC1-2)	33	6	20.80%
4	At5g13430	Q9LYR3	Ubiquinol--cytochrome-c reductase (REISKE subunit)	26	6	22.10%
5	At4g32470	Q9SUU5	Ubiquinol-cytochrome C reductase complex 14 kDa protein	14.5[d]	5	49.20%
	At5g25450	Q3E953	Ubiquinol-cytochrome C reductase complex 14 kDa protein	14.6[d]	1[d](2)	18.00%

1 **Acc. #:** Q0WW81 **Species:** ARATH **Name:** Hypothetical protein

Protein MW: 53326.2 **Protein pI:** 5.7

Num Unique	% Cov	Best Disc Score
24	41.1	3.25

M+H	m/z	z	Da	Peptide	S	Score	Score Diff	Disc Score	# in DB
872.7927	436.9000	2	0.14	VEIAELAK	351	24.5	11.9	1.95	2
897.2727	449.1400	2	0.42	DIADFTSK	345	26.3	16.6	2.47	3
903.8927	452.4500	2	0.19	SVDQLTLK	344	25.1	9.7	1.71	2
916.8927	458.9500	2	0.19	ELKDVAGGK	291	17.7	8.1	1.54	2
952.6727	47								
963.1527	48								
974.8754	32								
975.0927	48								
1028.8327	51								
1066.0454	35								
1120.0727	56								
1132.0527	56								
1136.1127	56								
1148.3054	38								
1198.1927	59								
1299.0727	65								
1315.4127	65								
1462.9127	73								
1626.1727	81								
1642.0327	82								
1732.1727	86								
1861.4654	62								
1986.4927	993.7500	2	0.20	VAEPLTSDLPNVPPQLAPK	389	28.1	21.0	2.95	2
2251.1054	751.0400	3	0.30	AALTSLDmPLQGVSLPPPLADK	401	23.2	9.1	1.65	2

1-1 **Acc. #:** O04308 **Species:** ARATH **Name**

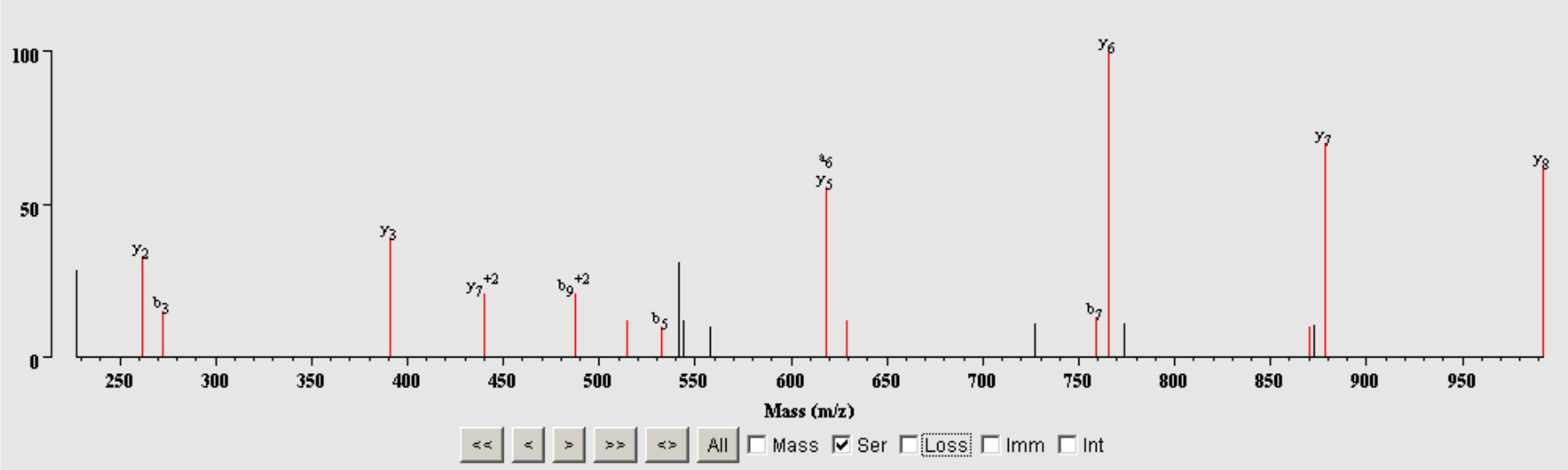
Protein MW: 54053.9 **Protein pI:** 6.0

Num Unique	% Cov	Best Disc Score
9	16.4	2.71

M+H	m/z	z	Da	Peptide	S	Score	Score Diff	Disc Score	# in DB
897.2727	449.1400	2	0.42	DIADFTSK	345	26.3	16.6	2.20	3
963.1527	482.0800	2	0.33	qILTYGER	367	22.8	13.2	1.82	3
974.8754	325.6300	3	0.10	KPVDQFLK	339	25.2	7.7	1.22	3
975.0927	488.0500	2	-0.20	MIAAEDIGR	341	20.5	11.3	1.61	3
1028.8327	514.9200	2	0.18	SQYVGGDFR	332	23.6	15.2	2.04	3
1149.9927	575.5000	2	0.20	SAILmNLESR	362	20.0	11.9	1.68	1
1462.9127	731.9600	2	0.100	EIEAIGGNTSASASR	334	30.7	21.3	2.71	3
1732.1727	866.5900	2	0.17	NPAFLDWEVNEELR	402	24.1	13.8	1.89	3
1861.4654	621.1600	3	0.51	NPAFLDWEVNEELRK	390	16.5	3.8	0.79	3

Parameters

Sequence: **SAILmNLESR**
Amino Acid Composition: **A1 E1 I1 L2 N1 R1 S2 m1**
Peptide N terminus: **Hydrogen**
Peptide C terminus: **Free Acid**
Constant Modification: **Amidomethyl Cysteine**
Instrument Name: **LCQ**
Ion Types Considered: **a, b, b-NH3, b-H2O, bp2, bp2-NH3, bp2-H2O, bp2-H3PO4, bp2-SOCH4, y, y-NH3, y-H2O, yp2, yp2-NH3, yp2-H2O, yp2-H3PO4, yp2-SO(**
All fragment ion masses below are calculated as: **monoisotopic** masses

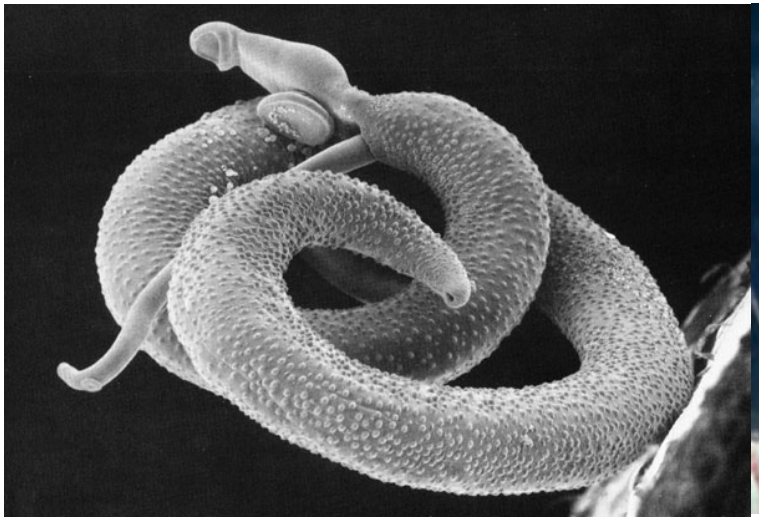


Max Intensity: **47039**
Num Matched: **14**
Num Peaks: **21**
Percent Unmatched: **33.3**

Schistosoma mansoni proteomikája

G. Knudsen, J.H McKerrow, UCSF

* kb. 200 millió ember fertőződik évente, bőrön keresztül, fertőzött vízben
* Átmeneti hordozó – *Biomphalaria glabrata*



* Kulcs a fertőzéshez:
  a cercaria által kiválasztott enzim-keverék

Egyszerű, jól ismert keverékkel van dolgunk...

- a csigából kibújt féreg-növendékeket jól lemossuk
- bőr-lipidekkel megindítjuk a kiválasztást
- a fehérje-oldatot bekoncentráljuk
- 1D SDS-PAGE fehérje-frakcionálásra
- gélben emésztés tripszinnel
- LC/MS/MS analízis

Eksigent/Famous - 75 mm ID PepMap column; 300 nL/min flow; water/ACN/0.1% HCOOH mobile phase, gradient elution,

QSTAR; 1 sec MS/3 sec CID; single charged and dynamic exclusion

Mit találtunk?

- cercaria felszíni fehérjéket
- cercaria elasztázt és glikolitikus enzimeket
- csiga immunválasz-fehérjéket
- fotoszintézisben résztvevő fehérjéket
- humán keratint
- marha szérum albumint

Honnan jött mindez?

- a cercaria külsejéről
- ez az igazi! – a kiválasztott fehérjék
- az átmeneti hordozóból
- a salátából, amit a csigák ettek – éheztessük őket?
- a technikus kezéről? hajából?
- a laborból mindenütt – lehet, hogy takarítani kellene?

Új, hatásosabb izolálási módszer→ az elasztáz-aktivitás 40x-esére nőtt!

bioinformatika = bio + informatika

- Nem szabad elfelejtenünk, hogy biológiai mintákkal dolgozunk!
- és kémiailag is aktív vegyületekkel!
- Az algoritmus lehet tökéletes, csak a biológiai rendszer nem viselkedik!

“Kémiai szennyezések” - ízelítő

-34 Da (-H_2S) – Cys

-17 Da (ciklizáció) – N-terminális Gln, carbamidometil Cys

+1 Da (amid hidrolízise) Gln, Asn

+14 Da (Me) – Glu, (Asp)

+16 Da (“O”) – Cys, Trp, Met

+32 Da (O_2) – Cys, Trp, Met

+43 Da (NH_2-CO-) – N-terminus, Lys

+48 Da (O_3) – Cys

+57 Da (-CH_2CONH_2) – N-terminus, Met, His, Asp, Lys

+71 Da (-CH_2-CH_2-$CONH_2$ = akrilamid) – Cys

Kovalens módosítások detektálása

- szerencsés esetben találhatunk poszt-transzlációs módosításokat automatizált MS-MS analízissel

- lekereső programok lehetővé teszik a kovalens módosítások követését

MASCOT MS/MS Ions Search

Your name		Email	
Search title	C:\PE Sciex Data\Projects\API Instrument\Data\L4072812.wiff : Sm CS		
Database	MSDB		
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)	Variable modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)
Protein mass	kDa	ICAT	☐
Peptide tol. ±	2.0 Da	MS/MS tol. ±	0.8 Da
Peptide charge	2+	Monoisotopic	◉ Average ○
Data file	1\LOCALS~1\Temp\mas23.tmp Browse...		
Data format	Mascot generic	Precursor	m/z
Instrument	Default		
Overview	☐	Report top	20 hits
	Start Search ...		Reset Form

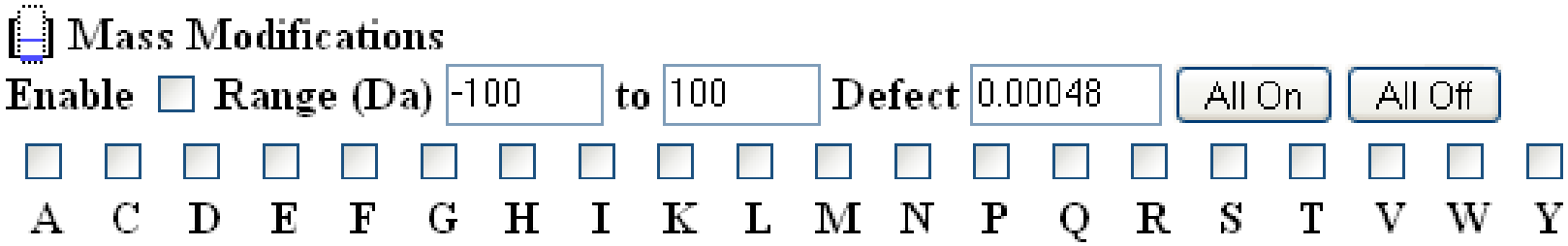
Túl sok variációs lehetőség rengeteg félreértelmezést okozhat !

Van azért kiskapu

Kiskapu 1

Unanticipated Modifications with ProteinProspector!

1. Search data with strict parameters
2. Search data again, but against only confidently identified proteins



The search speed is fast enough to analyze large datasets:

A search of 58,000 spectra considering any mass shift between –100Da and +200 Da on 100 proteins completed overnight on a dual processor computer.

Results output

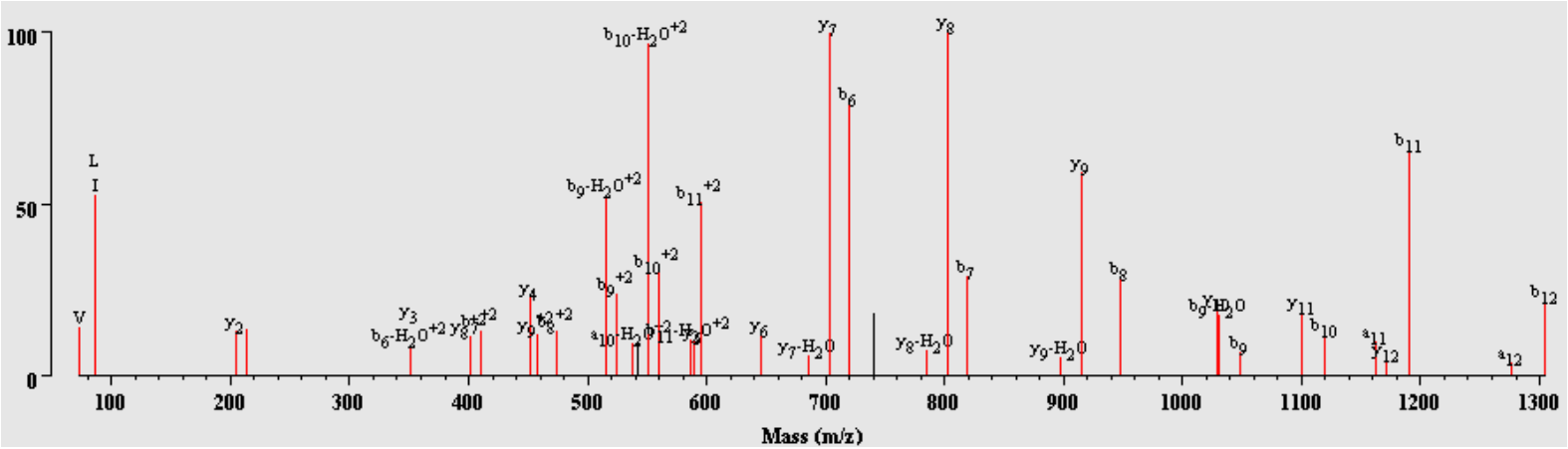
m/z	z	Da	Peptide	Expect
895.9686	2	0.0053	**K(26.0125)AALEQDLAFWYGPR**	1.7e-9
882.9565	2	-4.9e-4	**KAALEQDLAFWYGPR**	2.7e-9
1147.5895	2	-9.9e-4	**LVmASLYHIYVALEEEIER**	4.2e-9
984.5237	3	-1.9e-4	**IAQK(57.0274)ALDLPSSGEGLAFFTFPNIASATK**	9.3e-9
884.9588	2	8.1e-4	**KAALEQDLAFW(4.0019)YGPR**	2.2e-8
909.4724	3	0.34	**DGFKLVMASLYHIYVALEEEIER**	4.2e-8
1227.6336	2	0.0024	**ALDLPSSGEGLAFFTFPNIASATK**	6.0e-8
890.9590	2	-0.0019	**KAALEQDLAFW(16.0077)YGPR**	6.4e-8
898.9534	2	-0.011	**KAALEQDLAFW(32.0154)YGPR**	1.4e-7
834.9070	2	-0.010	**AALEQDLAFW(32.0154)YGPR**	1.4e-7
818.9078	2	-0.0017	**AALEQDLAFWYGPR**	1.7e-7

Mostly the “usual suspects”

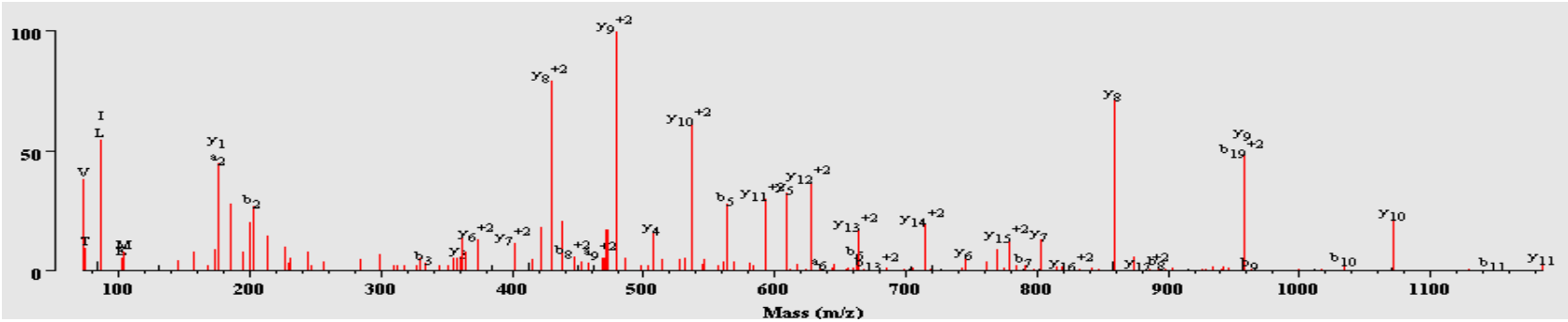
* Metal Adducts (+22, +38, +53)
* Oxidation and dioxidation of Trp (+16, +32)
* Trp oxidation to kynurenin (+4)
* Propionamide-modified Cys (+71)
* Deamidation of Asn (+1)
* Iodination of Tyr (+126)
* Side-chain loss of Cys (Cys->Dha)
* Carbamidomethylation of Met (+57)

Enzyme Catalyzed Rearrangement Reaction

RMAMNDVETAALIVGGHTFGK^{+3}



MAMNDVETAALIVGGHTFGK(156.0749)$^{+4}$



Lippincott, J. *et al.* (1997) *Anal. Biochem.* **252** p.314-325.

Spectral pairs: outline

Pre-processing

Clustering: combine multiple spectra from the same peptide

Protein identification

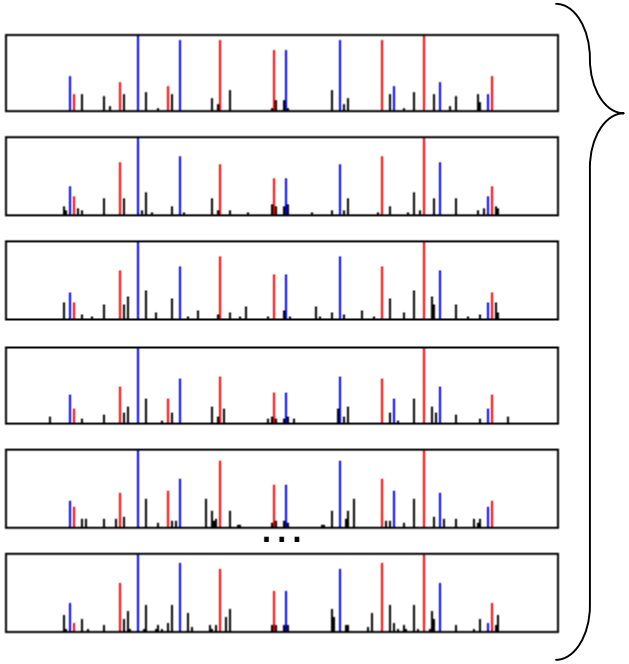
a. Find spectral pairs – spectra from peptide pairs.
b. Use spectral pairs to deconvolve prefix/suffix masses.
c. Generate a small set of long tags from deconvolved spectra.
d. Perform text search of protein sequences.

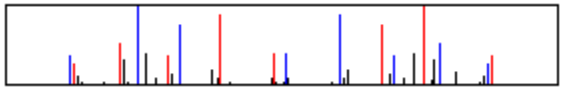
Identification of modifications

a. Single: non-amino acid mass differences between spectral pairs
b. Multiple: chains of spectral pairs

Preprocessing: clustering

Combines Bandeira et al.'04 and Ben-Dor et al.'99 with some scoring improvements.





Cluster consensus spectrum

	b	y	satellite	unexplained	signal to noise	
# masses	9.5	9.3	20.1	**35.3**	**0.27**	Single spectra
% intensity	28%	28%	19%	**25%**		
# masses	9.5	9.4	10.4	**13.7**	**0.69**	Cluster consensus spectra
% intensity	37%	36%	13%	**14%**		

Selection of relevant masses – averages over all single spectra and cluster consensus spectra

Modifications on cataractous lens

Location	Modification Mass	Putative annotation
S,T	-18	dehydration
Q	-17	deamidation
W	-2	cross-linking
H	14	methylation
M,W	16	oxidation
S,H	28	double methylation
N-term	42	acetylation
N-term	43	carbamylation
K,non-terminal	43	carbamylation
W	44	carboxylation
R	55	unknown
K	58	carboxymethylation
K	72	carboxyethylation

Table 1: Rediscovered all modifications previously identified by blind database search except deamidation on N/Q (+1 Da).

Table 2: Identified 6 new modification events

Location	Modification mass	Type	Putative annotation	Comment
M	-48	Chem. artifact	loss of methane sulfenic acid	*reported on same site*
W	4	PTM	kynurenine	*reported in cataractous lenses*
S	30/73	unknown	unknown	
W	32	PTM	formylkynurenine	*reported in cataractous lenses*
N-term	57	unknown	carboxyamidomethylation	possible chem.artifact
N-term	229/271	unknown	unknown	