CENTRO DE GENÓMICA Y
PROTEÓMICA
UNIVERSIDAD COMPLUTENSE DE MADRID

UNIVERSIDAD
COMPLUTENSE
MADRID

IDENTIFICACIÓN DE PROTEÍNAS MEDIANTE ESPECTROMETRÍA DE MASAS.

ANÁLISIS BIONFORMÁTICO

Mª Luisa Hernáez

Unidad de
PROTEÓMICA
UCM

ESTRATEGIAS PROTEÓMICAS

Muestras complejas. Extractos proteicos: proteomas, subproteomas...

Proteómica Tradicional
Escuela Europea
en gel

Proteómica ¨Shotgun¨
Escuela Americana
En solución

-Separación de Proteínas (1D-2D)

-Digestión de cada proteína

-MS y MS/MS

-Digestión de las proteínas

-Separación cromatográfica de péptidos

-LC-MS/MS

Identificación de proteínas mediante huella peptídica (MS, PMF)

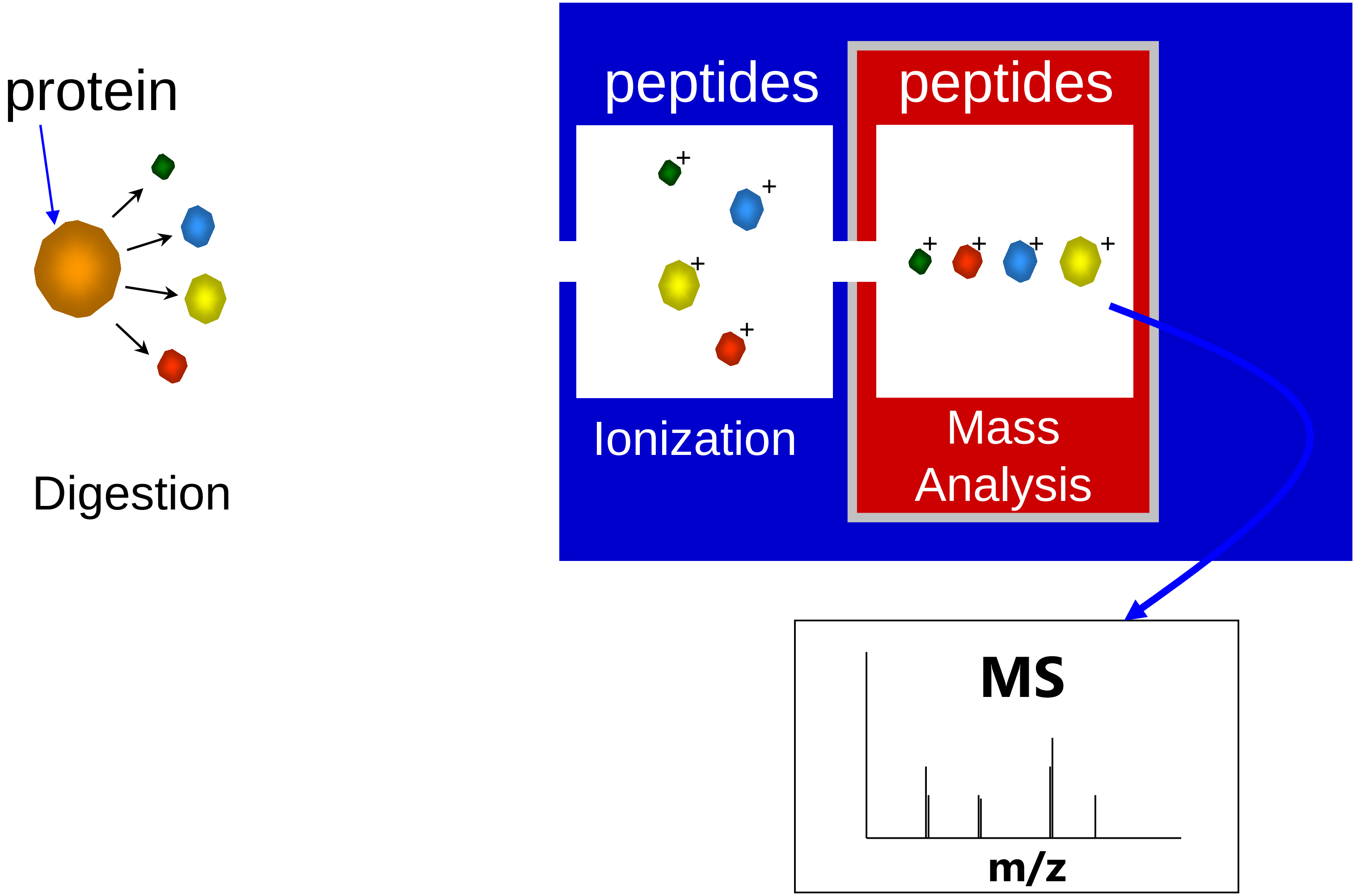
Espectrometría de masas

MALDI-TOF

Matrix-assisted laser desortion-ionization (MALDI).

Time of flight (TOF).

Single Stage Mass Spectrometry



protein

Digestion

peptides

Ionization

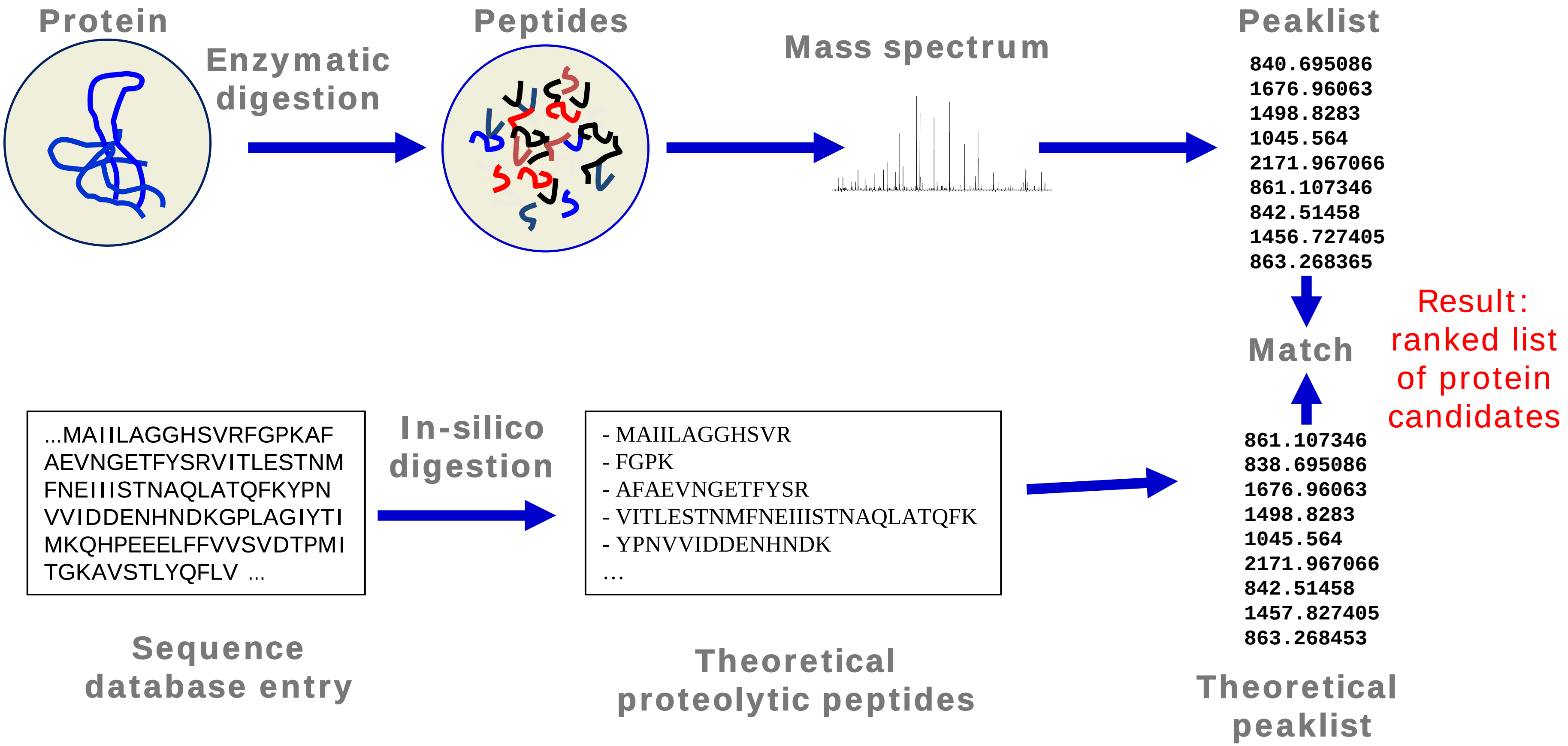
peptides

Mass Analysis

MS

m/z

Identificación de proteínas por huella peptídica



Protein

Enzymatic digestion

Peptides

Mass spectrum

Peaklist

840.695086
1676.96063
1498.8283
1045.564
2171.967066
861.107346
842.51458
1456.727405
863.268365

Match

Result: ranked list of protein candidates

...MAIILAGGHSVRFGPKAF
AEVNGETFYSRVITLESTNM
FNEIIISTNAQLATQFKYPN
VVIDDENHNDKGPLAGIYTI
MKQHPEEELFFVVSVDTPMI
TGKAVSTLYQFLV ...

Sequence database entry

In-silico digestion

- MAIILAGGHSVR
- FGPK
- AFAEVNGETFYSR
- VITLESTNMFNEIIISTNAQLATQFK
- YPNVVIDDENHNDK

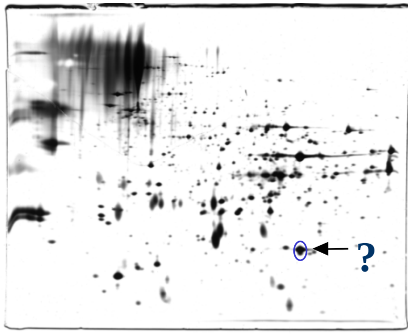
…

Theoretical proteolytic peptides

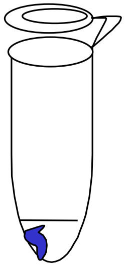
861.107346
838.695086
1676.96063
1498.8283
1045.564
2171.967066
842.51458
1457.827405
863.268453

Theoretical peaklist

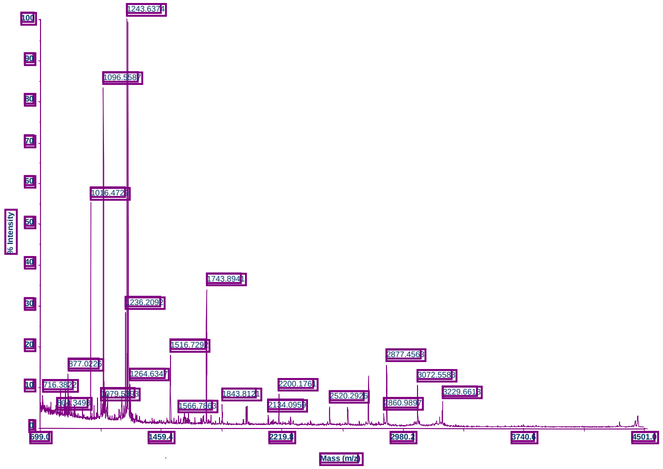
PROTEIN IDENTIFICATION BY MALDI-TOF MS



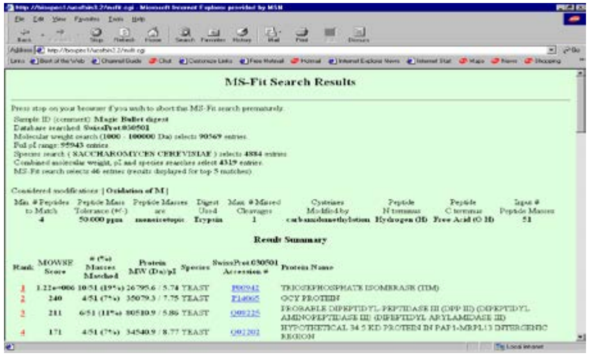
Separation of proteins by 2-D gel electrophoresis.
Excision of protein spots from 2 -D gel



In-gel trypsin digestion
Extraction of peptides mixture



Peptide mass fingerprint by
MALDI-TOF Mass Spectrometry



MS-Fit Search Results

Press stop on your browser if you wish to abort the MS-Fit search prematurely.

Sample ID (comment): **Magic Bullet digest**
Database searched: **SwissProt.030501**
Molecular weight search (**1000 - 100000** Da) selects **90569** entries.
Full pI range: **95943** entries.
Species search (**SACCHAROMYCES CEREVISIAE**) selects **4884** entries.
Combined molecular weight, pI and species searches select **4319** entries.
MS-Fit search selects 46 entries (results displayed for top 5 matches).

Considered modifications: [**Oxidation of M**]

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest Used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	**50.000 ppm**	**monoisotopic**	**Trypsin**	**1**	**carbamidomethylation**	**Hydrogen (H)**	**Free Acid (O H)**	**51**

Result Summary

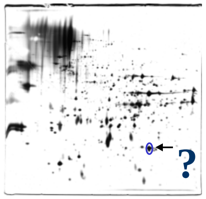
Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	SwissProt.030501 Accession #	Protein Name
1	1.22e+006	10/51 (19%)	26795.6 / 5.74	YEAST	P00942	TRIOSEPHOSPHATE ISOMERASE (TIM)
2	240	4/51 (7%)	35079.3 / 7.75	YEAST	P14065	GCY PROTEIN
3	211	6/51 (11%)	80510.9 / 5.86	YEAST	Q08225	PROBABLE DIPEPTIDYL-PEPTIDASE III (DPP III) (DIPEPTIDYL AMINOPEPTIDASE III) (DIPEPTIDYL ARYLAMIDASE III)
4	171	4/51 (7%)	34540.9 / 8.77	YEAST	Q03202	HYPOTHETICAL 34.5 KD PROTEIN IN PAP1-MRPL13 INTERGENIC REGION

Database searching

Protein identification:triose phosfate isomerase

Tratamiento de las manchas proteícas

Corte del gel



- Evitar contaminaciones
- Digerir la proteína eficientemente
- Recuperar la mayor cantidad de péptidos
- Minimizar las pérdidas por manipulaciones

Destinción

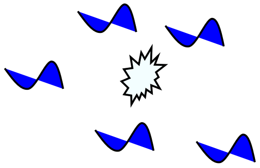
DTT IODOACETAMIDA

Reducción y alquilación

Deshidratación

Hidratación con tripsina y digestión 37ºC O/N

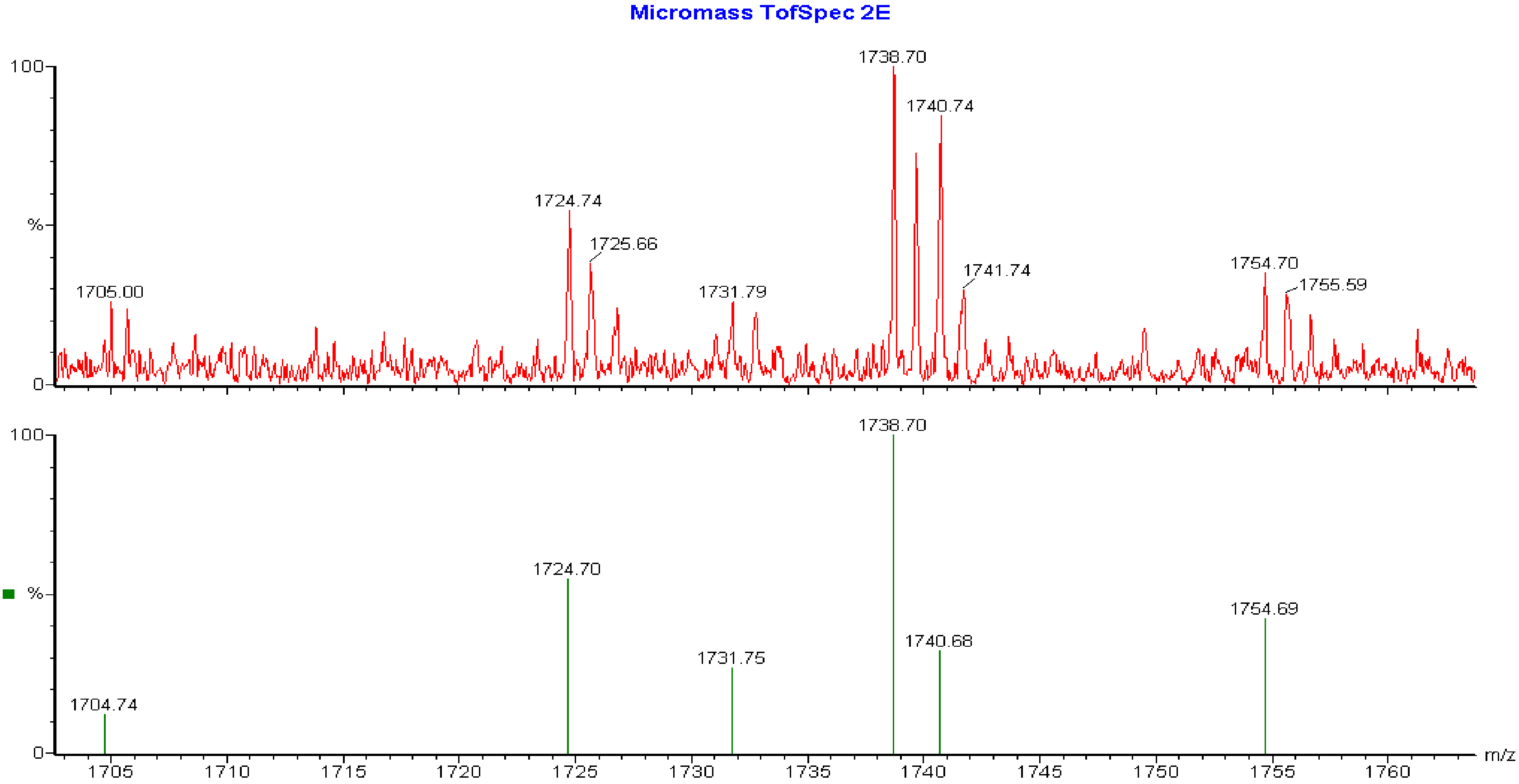
Extracción de los péptidos



¿Concentración y limpieza?

Espectro de MALDI

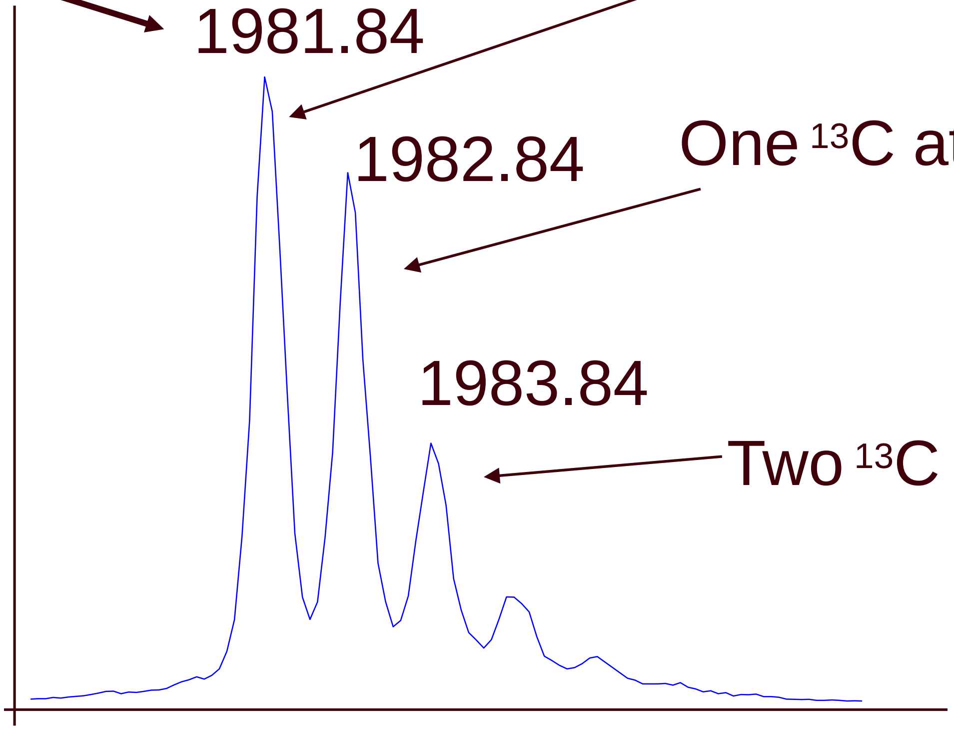
Abundancia relativa



Masa (m/z)

Resolución isotópica: Capacidad para separar dos picos que se diferencian en una unidad de masa.

“Masa monoisotópica”



1981.84

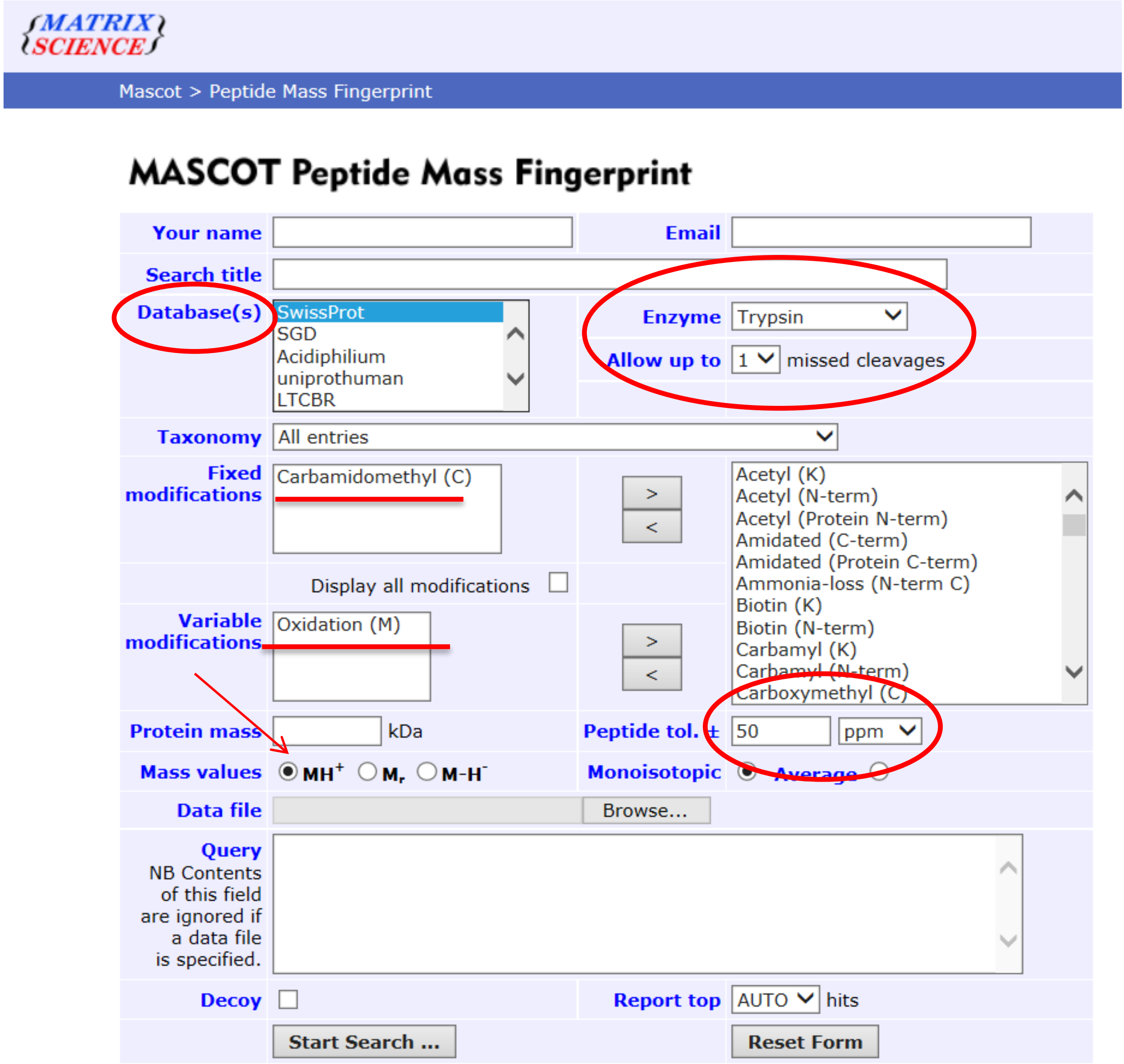
No ^{13}C atoms (all ^{12}C)

1982.84

One ^{13}C atom

1983.84

Two ^{13}C atoms



MASCOT Peptide Mass Fingerprint

Your name:
Email:
Search title:
Database(s): SwissProt (selected), SGD, Acidiphilium, uniprothuman, LTCBR
Enzyme: Trypsin
Allow up to 1 missed cleavages
Taxonomy: All entries
Fixed modifications: Carbamidomethyl (C)
Display all modifications ☐
Variable modifications: Oxidation (M)
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamyl (K)
Carbamyl (N-term)
Carboxymethyl (C)
Protein mass ___ kDa
Peptide tol. ± 50 ppm
Mass values ● MH$^+$ ○ M$_r$ ○ M-H$^-$
Monoisotopic ● Average ○
Data file
Query
NB Contents of this field are ignored if a data file is specified.
Decoy ☐
Report top AUTO hits

Tolerancia: es la diferencia entre el valor de la masa medida experimentalmente y el valor teórico

absoluta (Da) = [M exp-M teórica]

relativa (%, ppm) = [M exp-M teórica]/ M teórica

Análisis de péptidos trípticos

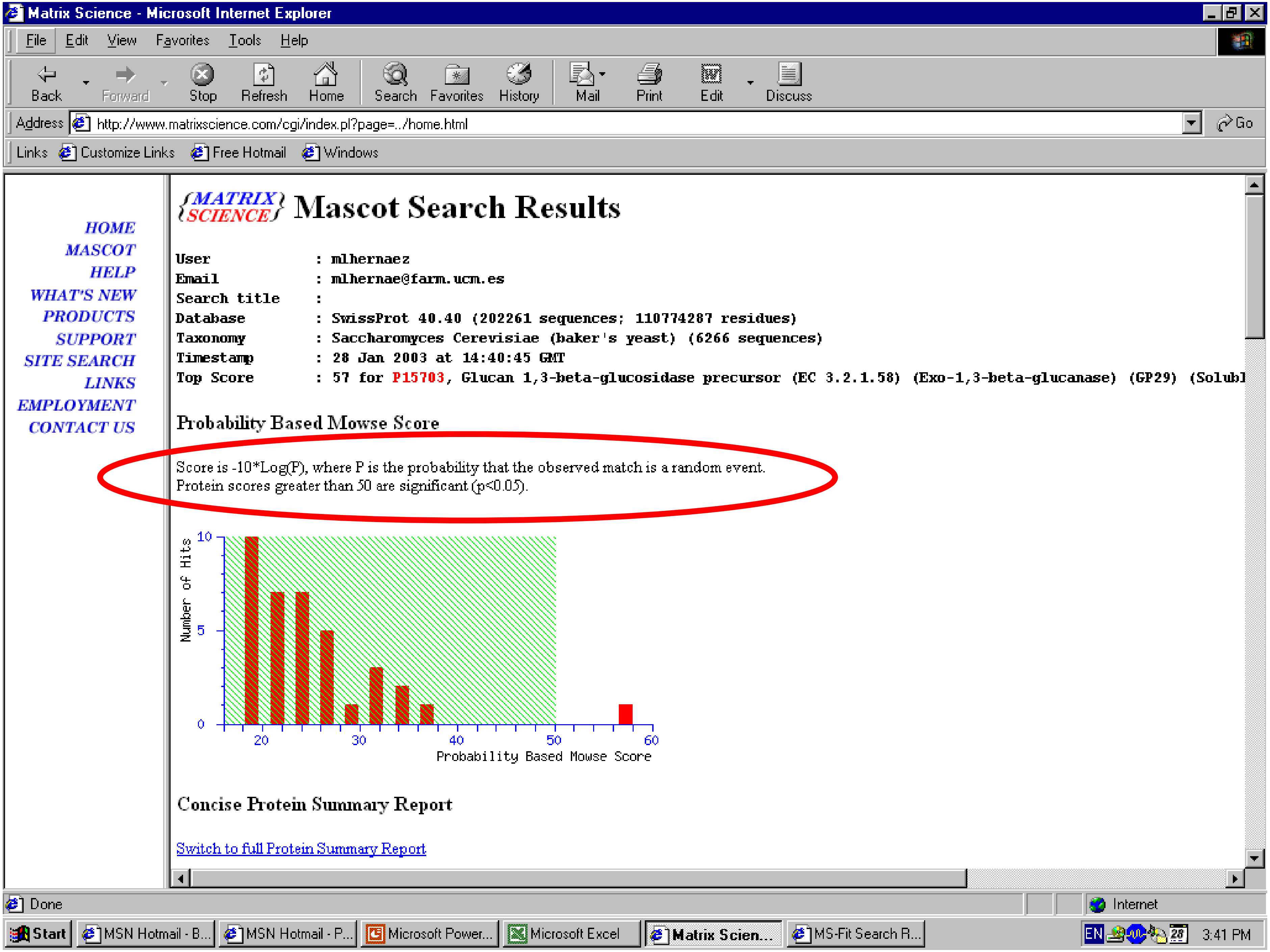
TRIPSINA: corte muy específico (lisina y arginina)

K-X, R-X excepto K-P, R-P

En proteínas de mamífero el contenido en Arg es 6% y en Lys es un 5%, (cada 100 aminoácidos hay 11 sitios de corte de tripsina)

Tamaño medio de los péptidos trípticos es de 9 aminoácidos: fácil extracción de péptidos del gel.

En este intervalo de masas (800 Da –2000 Da) la resolución es buena



Mascot Search Results

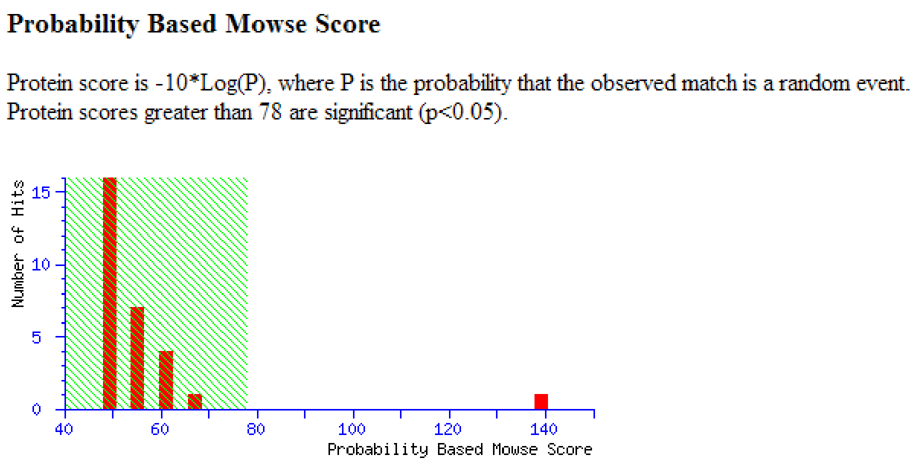
```
User           : mlhernaez
Email          : mlhernae@farm.ucm.es
Search title   :
Database       : SwissProt 40.40 (202261 sequences; 110774287 residues)
Taxonomy       : Saccharomyces Cerevisiae (baker's yeast) (6266 sequences)
Timestamp      : 28 Jan 2003 at 14:40:45 GMT
Top Score      : 57 for P15703, Glucan 1,3-beta-glucosidase precursor (EC 3.2.1.58) (Exo-1,3-beta-glucanase) (GP29) (Solubl
```

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 50 are significant (p<0.05).

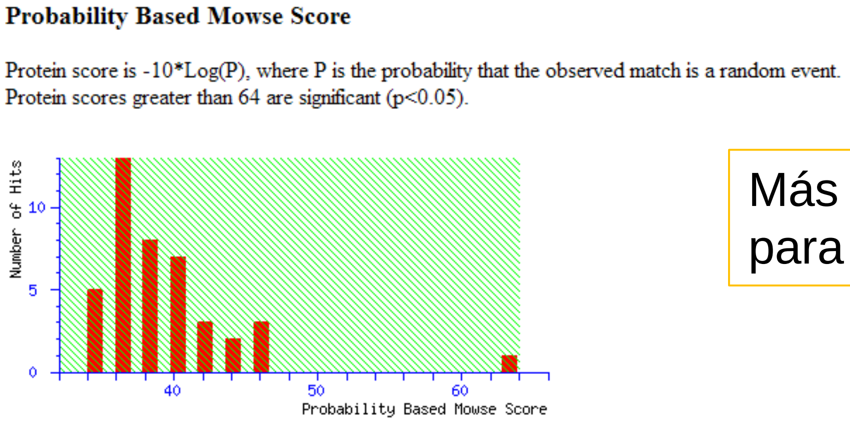
Concise Protein Summary Report

Switch to full Protein Summary Report



Probability Based Mowse Score

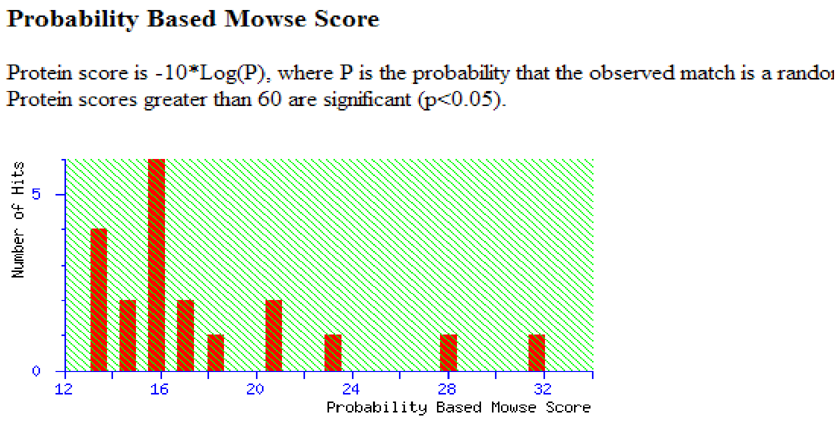
Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 78 are significant ($p<0.05$).



Probability Based Mowse Score

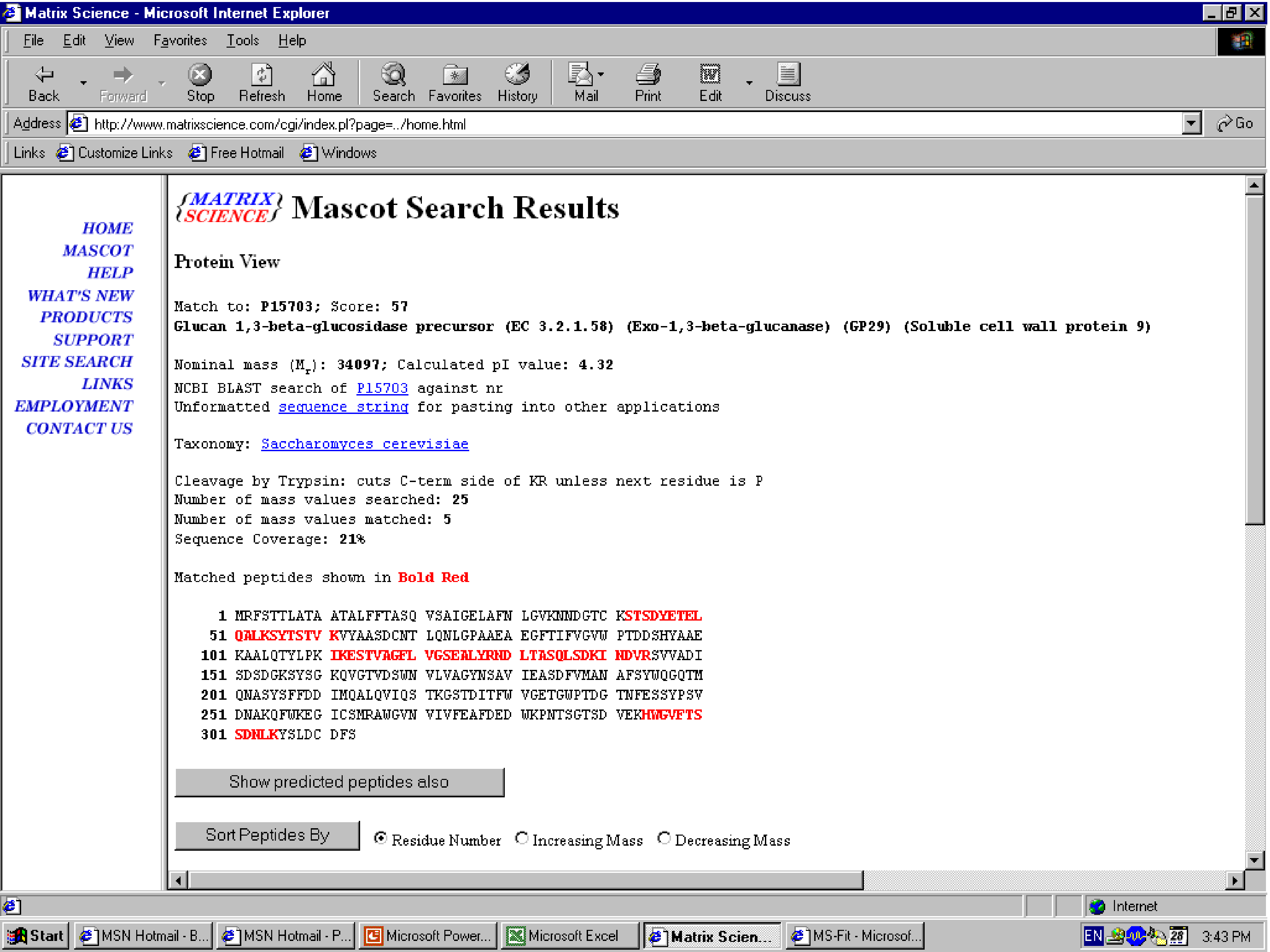
Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 64 are significant ($p<0.05$).

Más restrictivos en la búsqueda para que aumente la puntuación



Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a rando
Protein scores greater than 60 are significant ($p<0.05$).



Mascot Search Results

Protein View

Match to: **P15703**; Score: **57**
Glucan 1,3-beta-glucosidase precursor (EC 3.2.1.58) (Exo-1,3-beta-glucanase) (GP29) (Soluble cell wall protein 9)

Nominal mass (M_r): **34097**; Calculated pI value: **4.32**
NCBI BLAST search of P15703 against nr
Unformatted sequence string for pasting into other applications

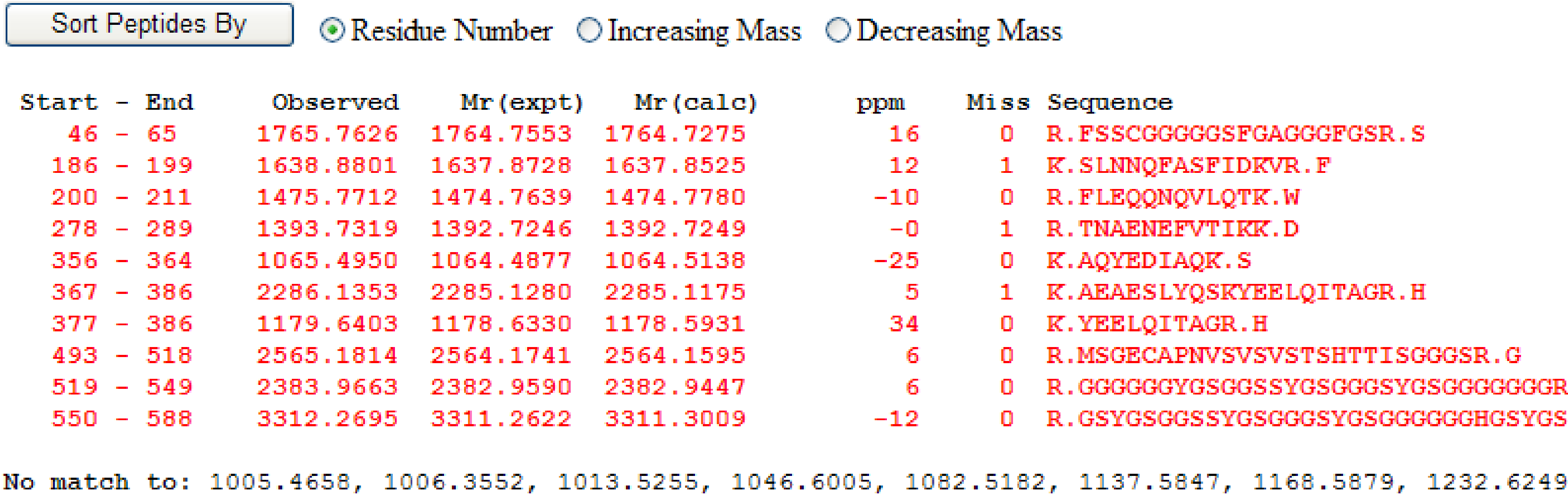
Taxonomy: Saccharomyces cerevisiae

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **25**
Number of mass values matched: **5**
Sequence Coverage: **21%**

Matched peptides shown in **Bold Red**

1 MRFSTTLATA ATALFFTASQ VSAIGELAFN LGVKNNDGTC K**STSDYETEL**
51 **QALKSYTSTV K**VYAASDCNT LQNLGPAAEA EGFTIFVGVW PTDDSHYAAE
101 KAALQTYLPK **IKESTVAGFL VGSEALYRND LTASQLSDKI NDVR**SVVADI
151 SDSDGKSYSG KQVGTVDSWN VLVAGYNSAV IEASDFVMAN AFSYWQGQTM
201 QNASYSFFDD IMQALQVIQS TKGSTDITFW VGETGWPTDG TNFESSYPSV
251 DNAKQFWKEG ICSMRAWGVN VIVFEAFDED WKPNTSGTSD VEK**HWGVFTS**
301 **SDNLK**YSLDC DFS

● Residue Number ○ Increasing Mass ○ Decreasing Mass



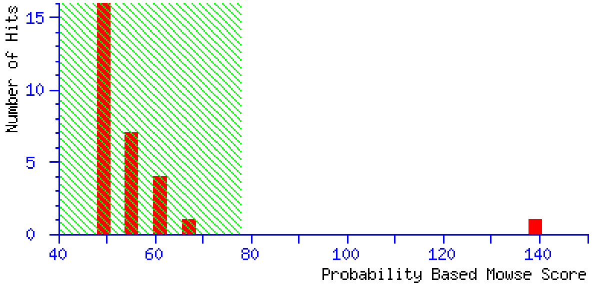
● Residue Number ○ Increasing Mass ○ Decreasing Mass

```
 Start - End     Observed   Mr(expt)   Mr(calc)      ppm    Miss Sequence
   46 - 65     1765.7626  1764.7553  1764.7275       16      0  R.FSSCGGGGGSFGAGGGFGSR.S
  186 - 199    1638.8801  1637.8728  1637.8525       12      1  K.SLNNQFASFIDKVR.F
  200 - 211    1475.7712  1474.7639  1474.7780      -10      0  R.FLEQQNQVLQTK.W
  278 - 289    1393.7319  1392.7246  1392.7249       -0      1  R.TNAENEFVTIKK.D
  356 - 364    1065.4950  1064.4877  1064.5138      -25      0  K.AQYEDIAQK.S
  367 - 386    2286.1353  2285.1280  2285.1175        5      1  K.AEAESLYQSKYEELQITAGR.H
  377 - 386    1179.6403  1178.6330  1178.5931       34      0  K.YEELQITAGR.H
  493 - 518    2565.1814  2564.1741  2564.1595        6      0  R.MSGECAPNVSVSVSTSHTTISGGGSR.G
  519 - 549    2383.9663  2382.9590  2382.9447        6      0  R.GGGGGGYGSGGSSYGSGGGSYGSGGGGGGGR
  550 - 588    3312.2695  3311.2622  3311.3009      -12      0  R.GSYGSGGSSYGSGGGSYGSGGGGGGHGSYGS

No match to: 1005.4658, 1006.3552, 1013.5255, 1046.6005, 1082.5182, 1137.5847, 1168.5879, 1232.6249
```

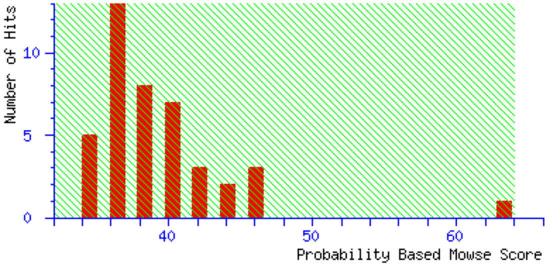
Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 78 are significant (p<0.05).



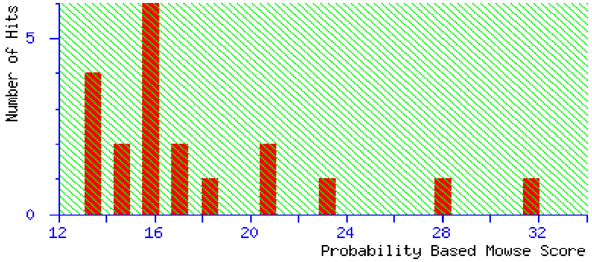
Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 64 are significant (p<0.05).



Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a rando
Protein scores greater than 60 are significant (p<0.05).



No hay éxito en la identificación?????

1. Número de péptidos en la huella peptídica es insuficiente para una coincidencia estadística (digestión, glicosilación)

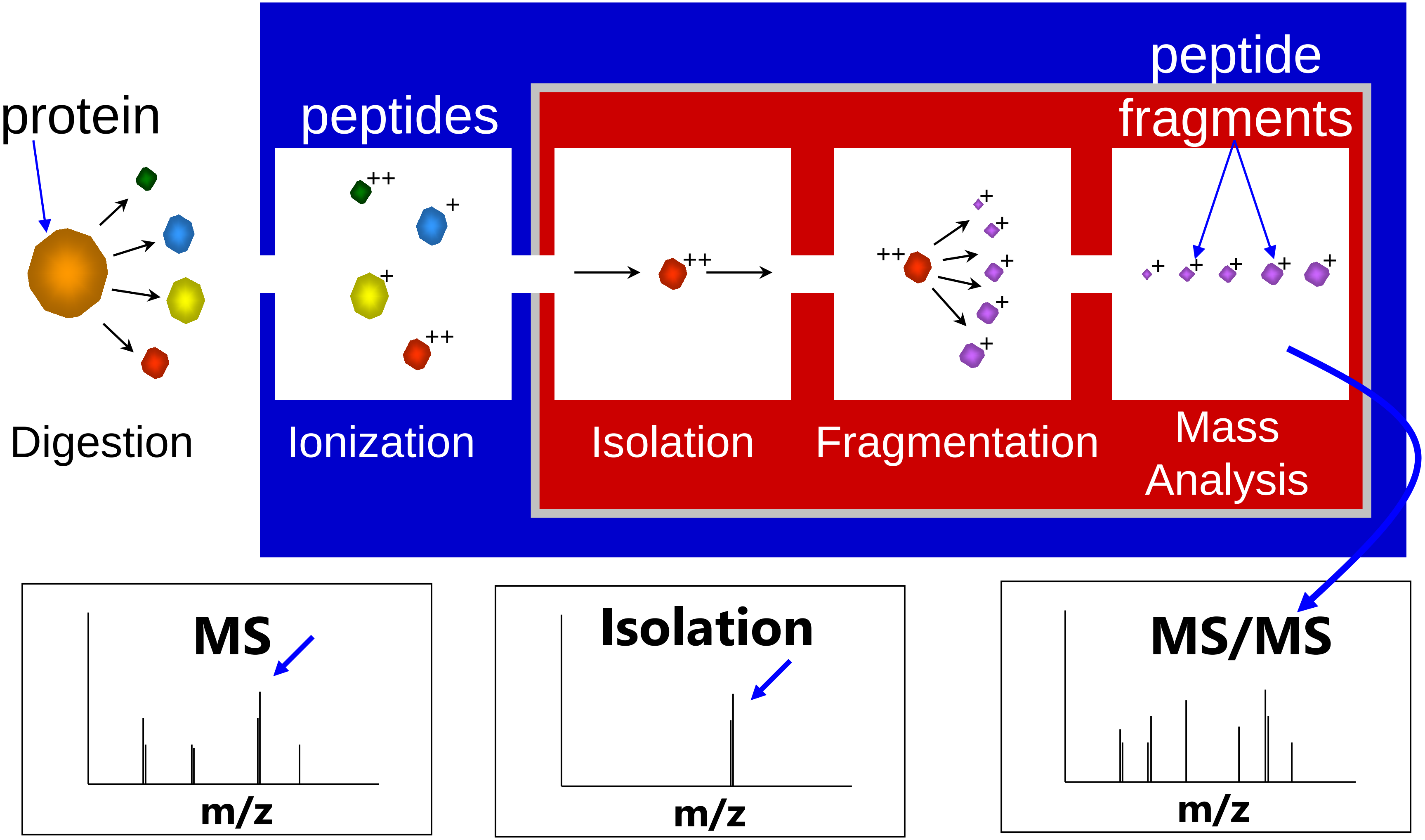
2. Mezcla de proteínas: supresión iónica, HPLC acoplado a MALDI-TOF

3. No está anotada en la base de datos de proteínas

Necesitamos más información

Tandem Mass Spectrometry

MS/MS usa 2 analizadores de masa (combinados en un instrumento) para seleccionar un analito (ion) de una mezcla, entonces genera fragmentos del ion aislado para obtener información estructural.



protein

peptides

peptide fragments

Digestion

Ionization

Isolation

Fragmentation

Mass Analysis

MS

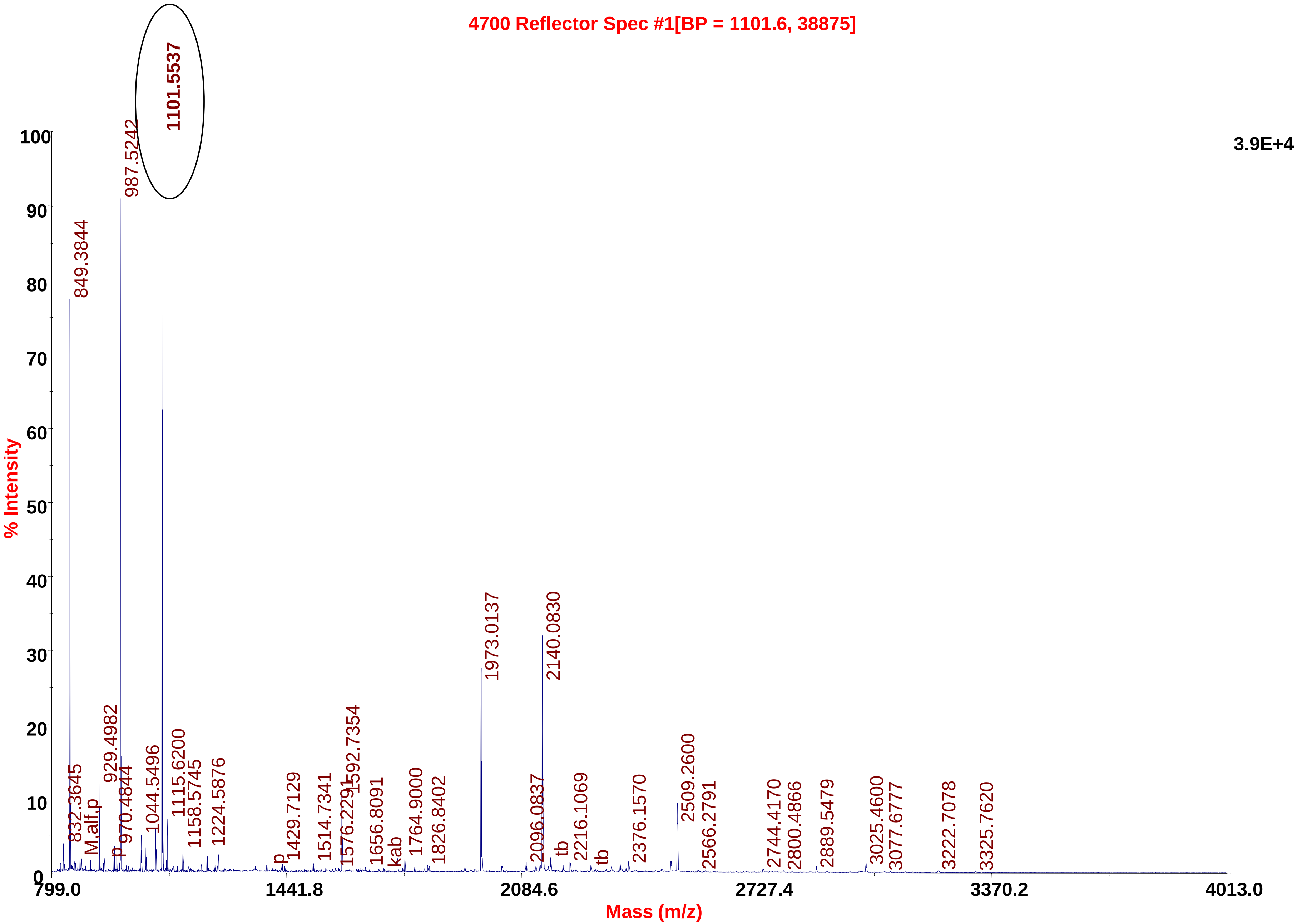
m/z

Isolation

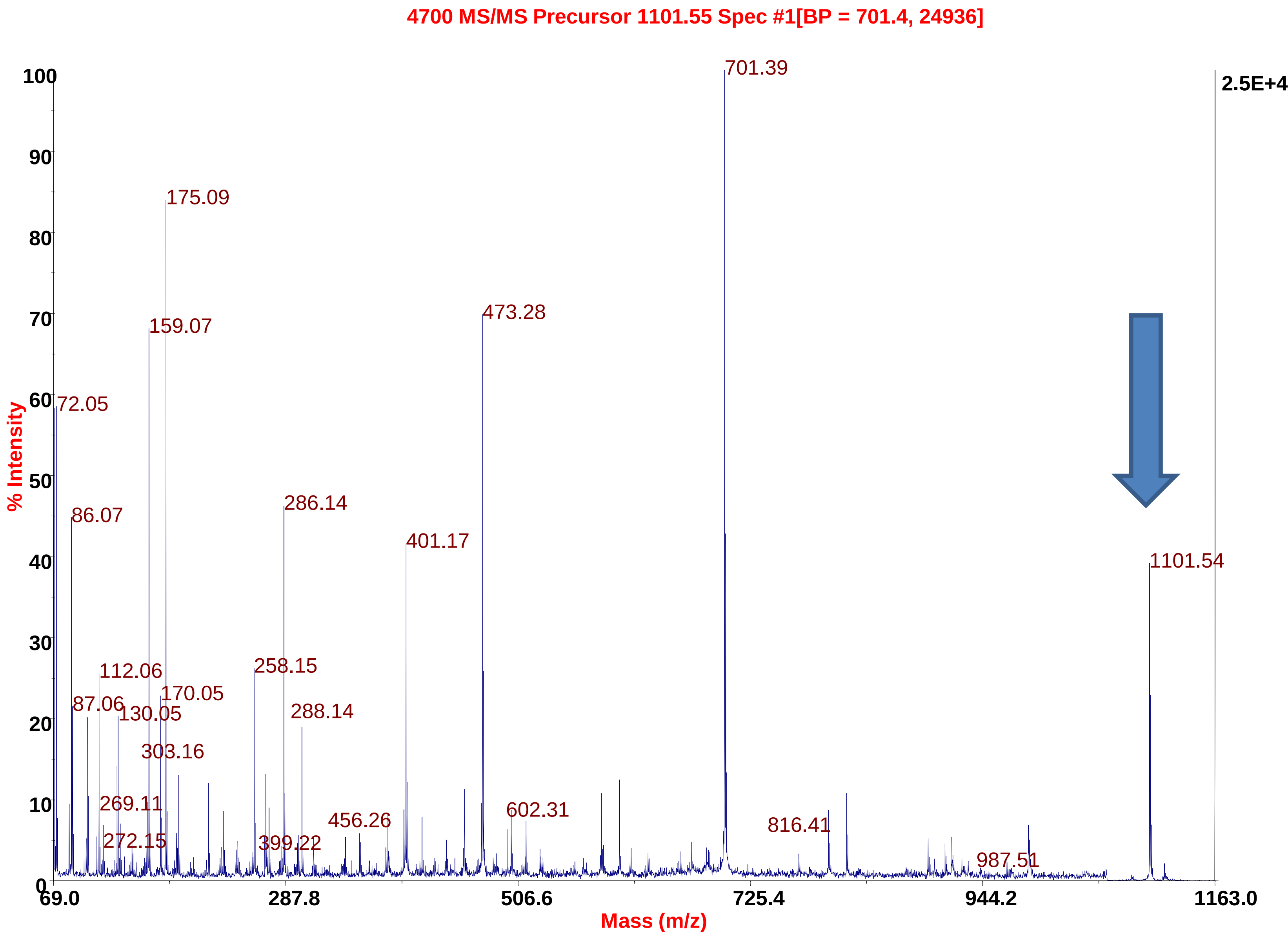
m/z

MS/MS

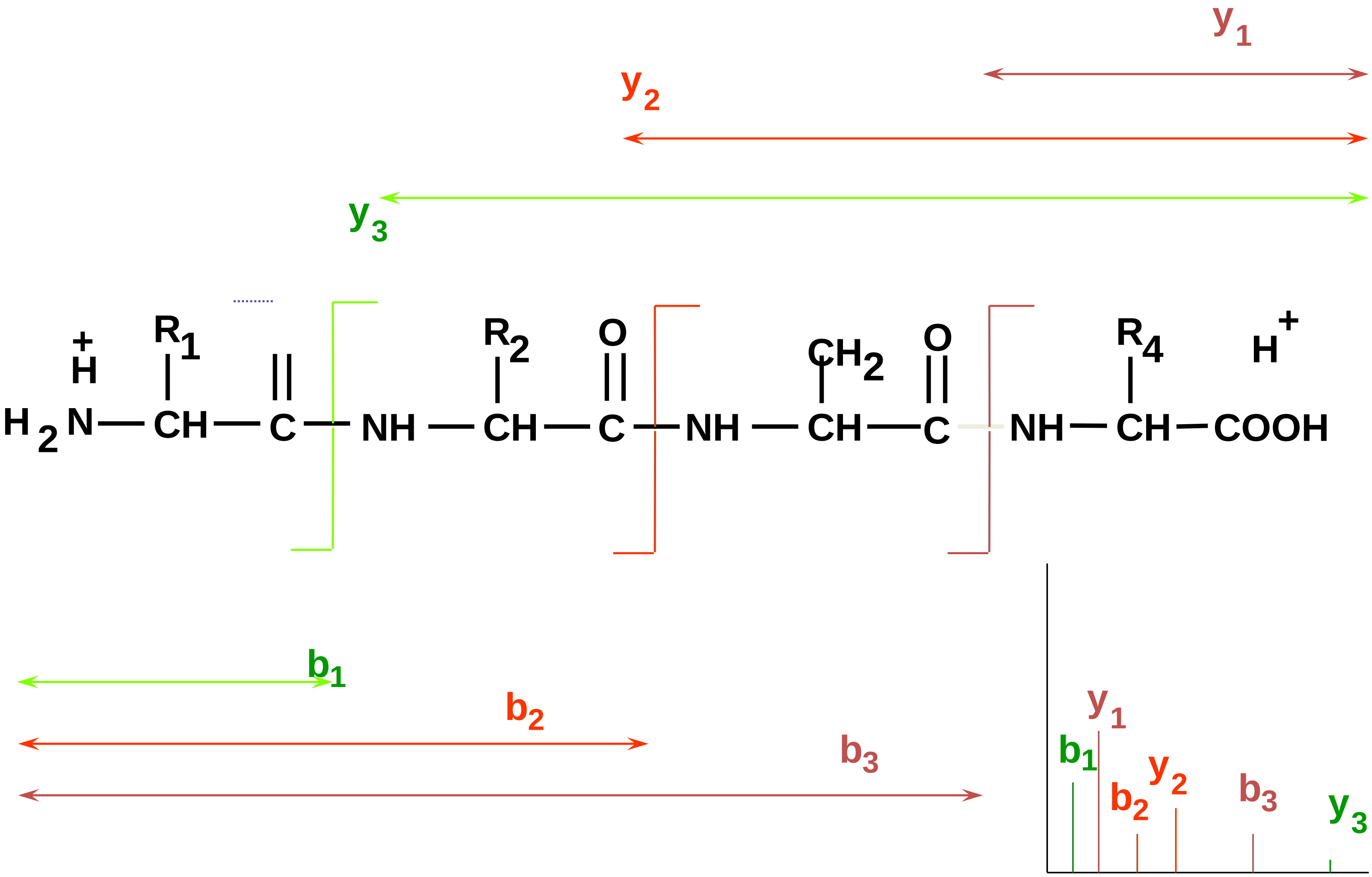
m/z


4700 Reflector Spec #1[BP = 1101.6, 38875]
% Intensity
Mass (m/z)
3.9E+4
100
90
80
70
60
50
40
30
20
10
0
799.0
1441.8
2084.6
2727.4
3370.2
4013.0
1101.5537
987.5242
849.3844
832.3645
M,alf,p
929.4982
p
970.4844
1044.5496
1115.6200
1158.5745
1224.5876
p
1429.7129
1514.7341
1576.2291
1592.7354
1656.8091
kab
1764.9000
1826.8402
1973.0137
2096.0837
2140.0830
tb
2216.1069
tb
2376.1570
2509.2600
2566.2791
2744.4170
2800.4866
2889.5479
3025.4600
3077.6777
3222.7078
3325.7620

Fragmentación del ión precursor:

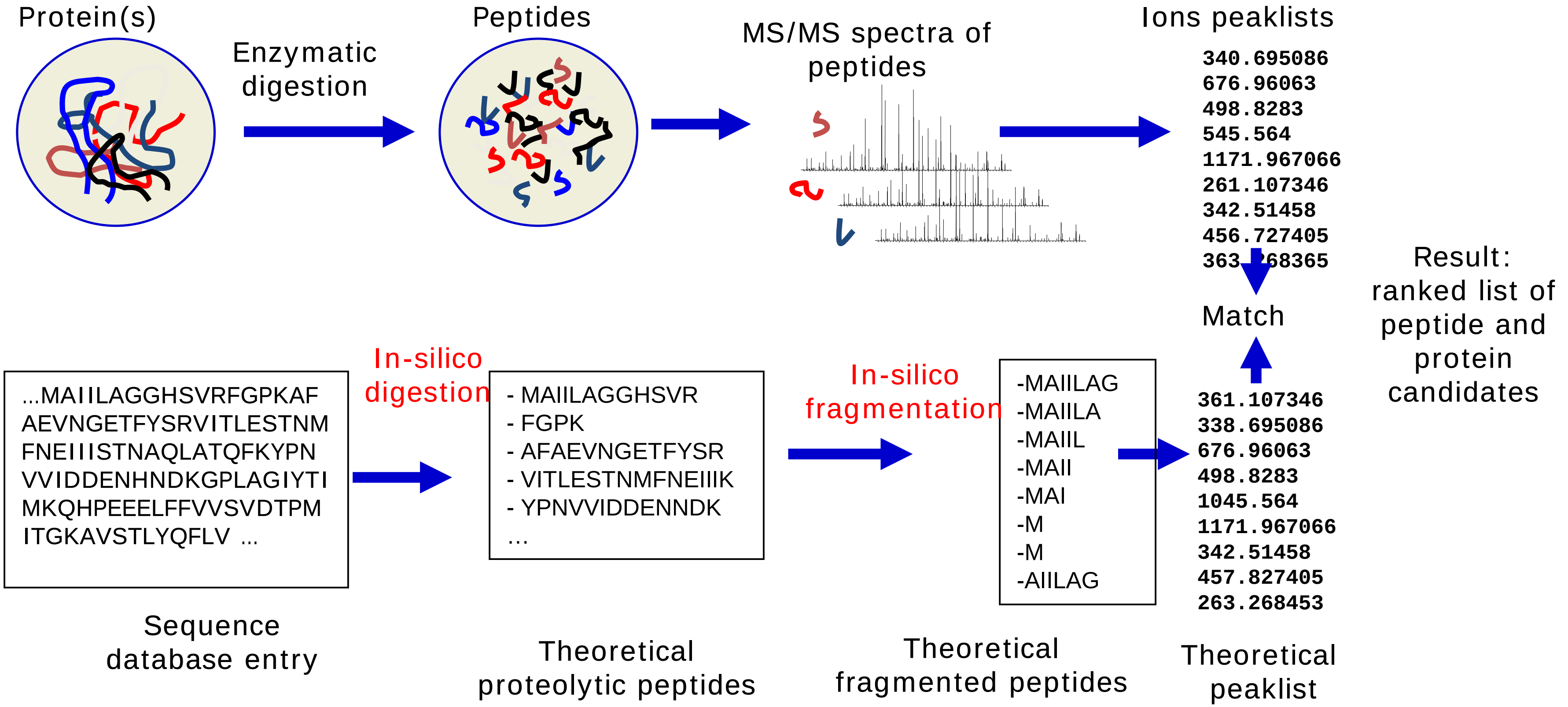


Los iones se fragmentan frecuentemente por los enlaces peptídicos dando lugar a una “escalera” de iones peptídicos característica de la secuencia del péptido. La diferencia de masa entre iones consecutivos define un residuos aminoacídico.



Peptide Identification by MS/MS

MS/MS database matching using fragment ion masses



Protein(s)

Enzymatic digestion

Peptides

MS/MS spectra of peptides

Ions peaklists

340.695086
676.96063
498.8283
545.564
1171.967066
261.107346
342.51458
456.727405
363.268365

Match

Result:
ranked list of peptide and protein candidates

...MAIILAGGHSVRFGPKAF
AEVNGETFYSRVITLESTNM
FNEIIISTNAQLATQFKYPN
VVIDDENHNDKGPLAGIYTI
MKQHPEEELFFVVSVDTPM
ITGKAVSTLYQFLV ...

Sequence database entry

In-silico digestion

- MAIILAGGHSVR
- FGPK
- AFAEVNGETFYSR
- VITLESTNMFNEIIIK
- YPNVVIDDENNDK

...

Theoretical proteolytic peptides

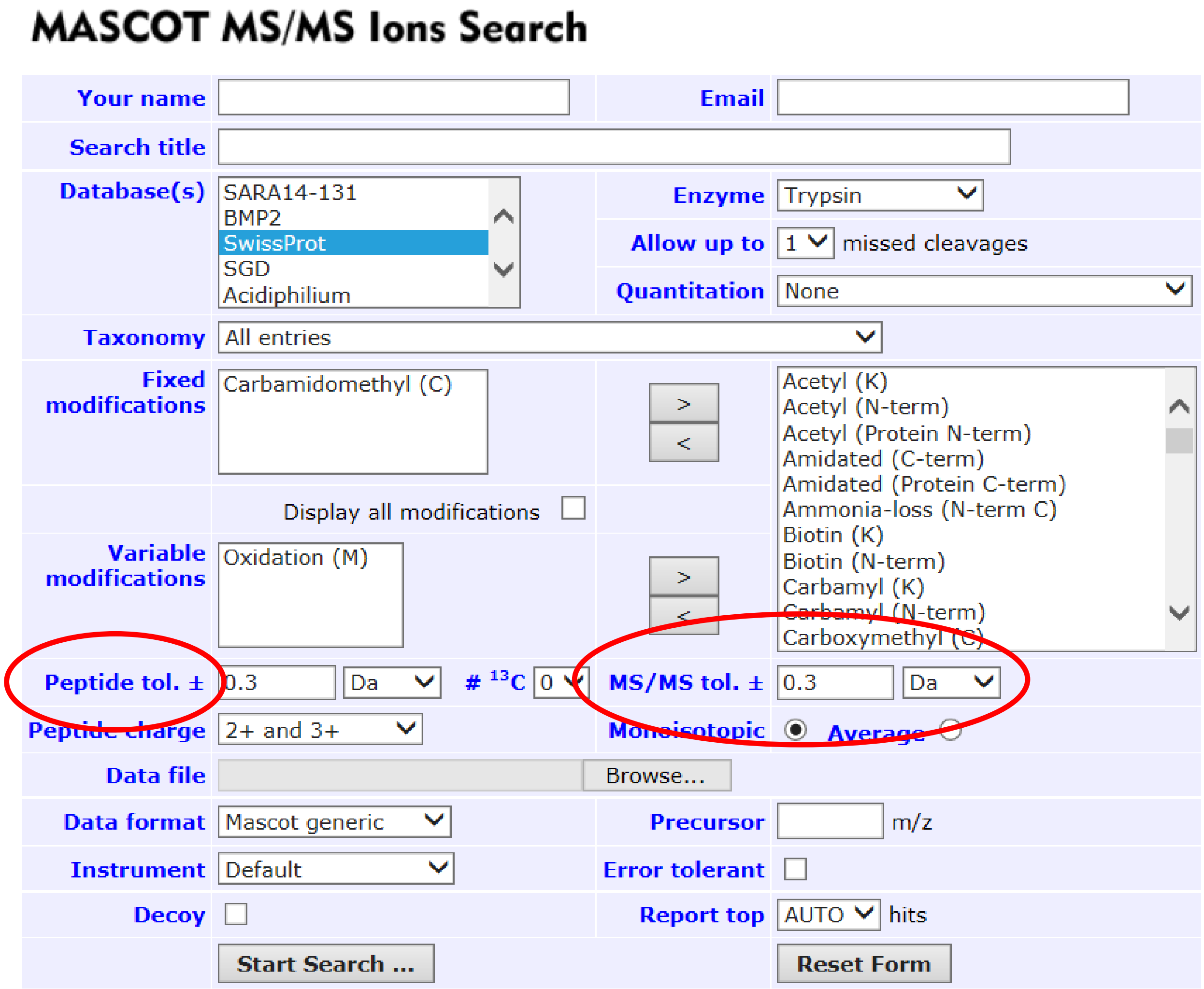
In-silico fragmentation

-MAIILAG
-MAIILA
-MAIIL
-MAII
-MAI
-M
-M
-AIILAG

Theoretical fragmented peptides

361.107346
338.695086
676.96063
498.8283
1045.564
1171.967066
342.51458
457.827405
263.268453

Theoretical peaklist



MASCOT MS/MS Ions Search

Your name
Email

Search title

Database(s) SARA14-131, BMP2, SwissProt, SGD, Acidiphilium
Enzyme Trypsin
Allow up to 1 missed cleavages
Quantitation None

Taxonomy All entries

Fixed modifications Carbamidomethyl (C)

Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamyl (K)
Carbamyl (N-term)
Carboxymethyl (C)

Display all modifications ☐

Variable modifications Oxidation (M)

Peptide tol. ± 0.3 Da **#** ^{13}**C** 0
MS/MS tol. ± 0.3 Da

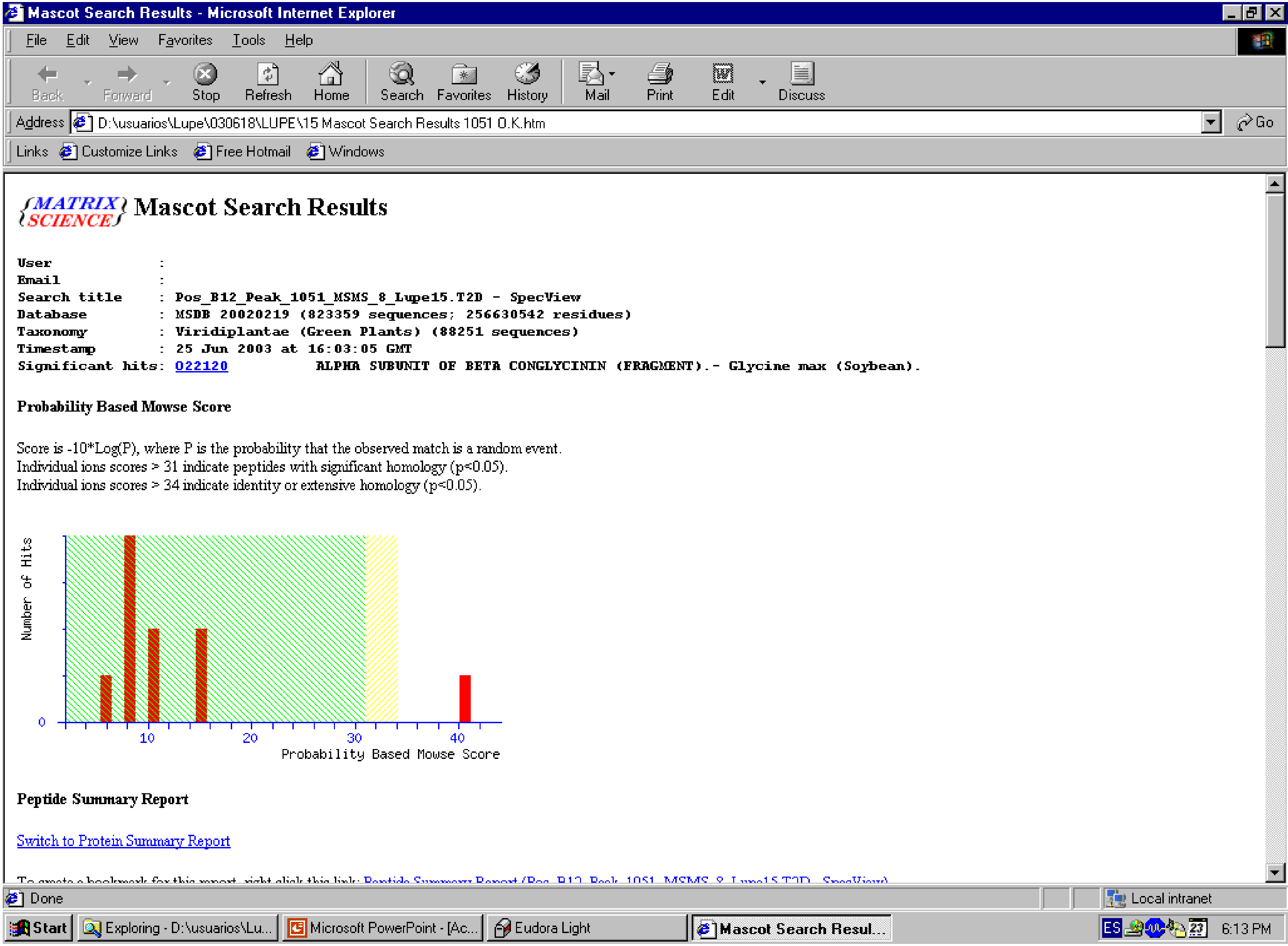
Peptide charge 2+ and 3+
Monoisotopic ● **Average** ○

Data file Browse...

Data format Mascot generic
Precursor m/z

Instrument Default
Error tolerant ☐

Decoy ☐
Report top AUTO hits



Mascot Search Results

User :
Email :
Search title : **Pos_B12_Peak_1051_MSMS_8_Lupe15.T2D - SpecView**
Database : **MSDB 20020219 (823359 sequences; 256630542 residues)**
Taxonomy : **Viridiplantae (Green Plants) (88251 sequences)**
Timestamp : **25 Jun 2003 at 16:03:05 GMT**
Significant hits: 022120 **ALPHA SUBUNIT OF BETA CONGLYCININ (FRAGMENT).- Glycine max (Soybean).**

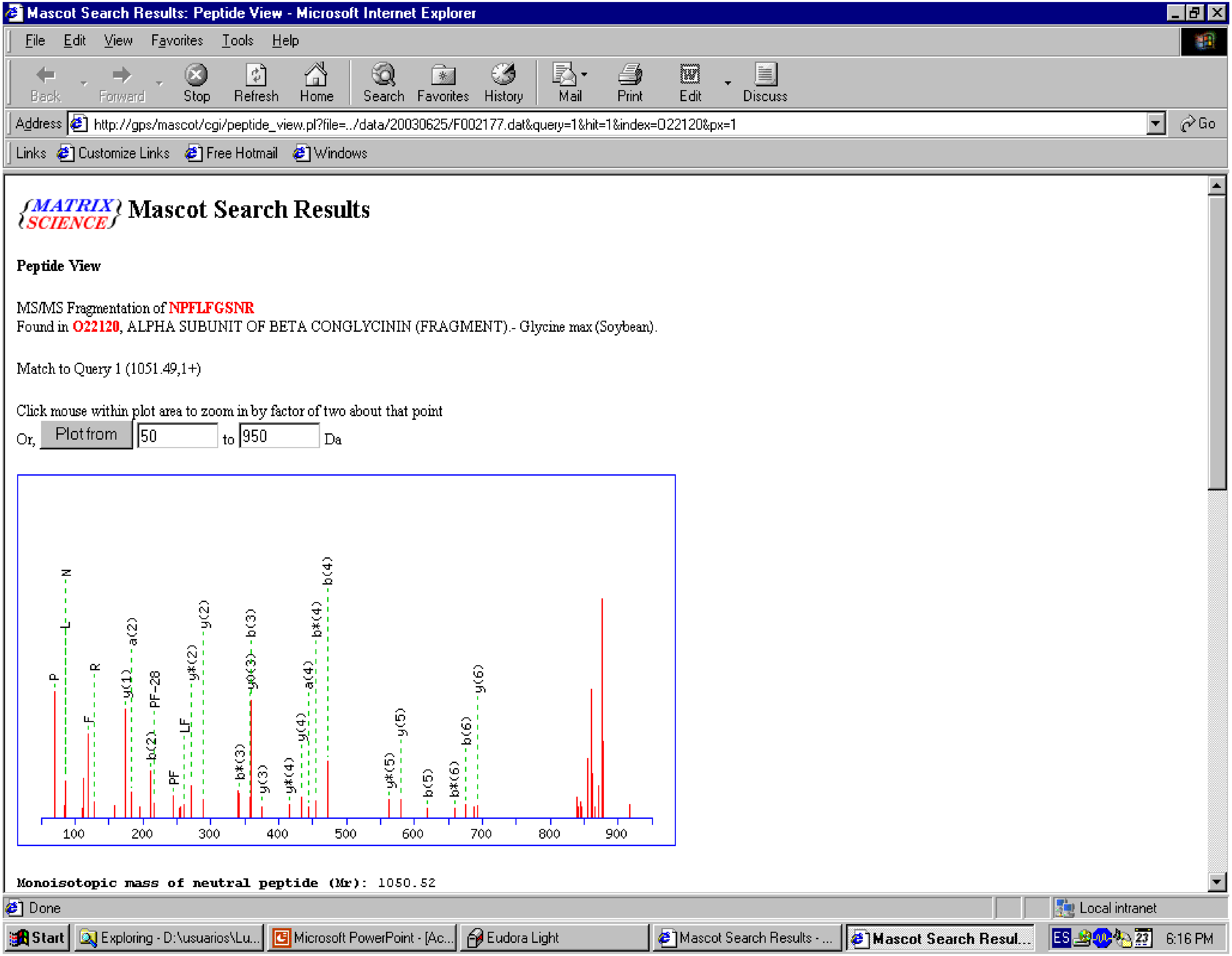
Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 31 indicate peptides with significant homology (p<0.05).
Individual ions scores > 34 indicate identity or extensive homology (p<0.05).

Peptide Summary Report

Switch to Protein Summary Report

To create a bookmark for this report, right click this link: Peptide Summary Report (Pos_B12_Peak_1051_MSMS_8_Lupe15.T2D - SpecView)



http://gps/mascot/cgi/peptide_view.pl?file=../data/20030625/F002177.dat&query=1&hit=1&index=O22120&px=1

Mascot Search Results

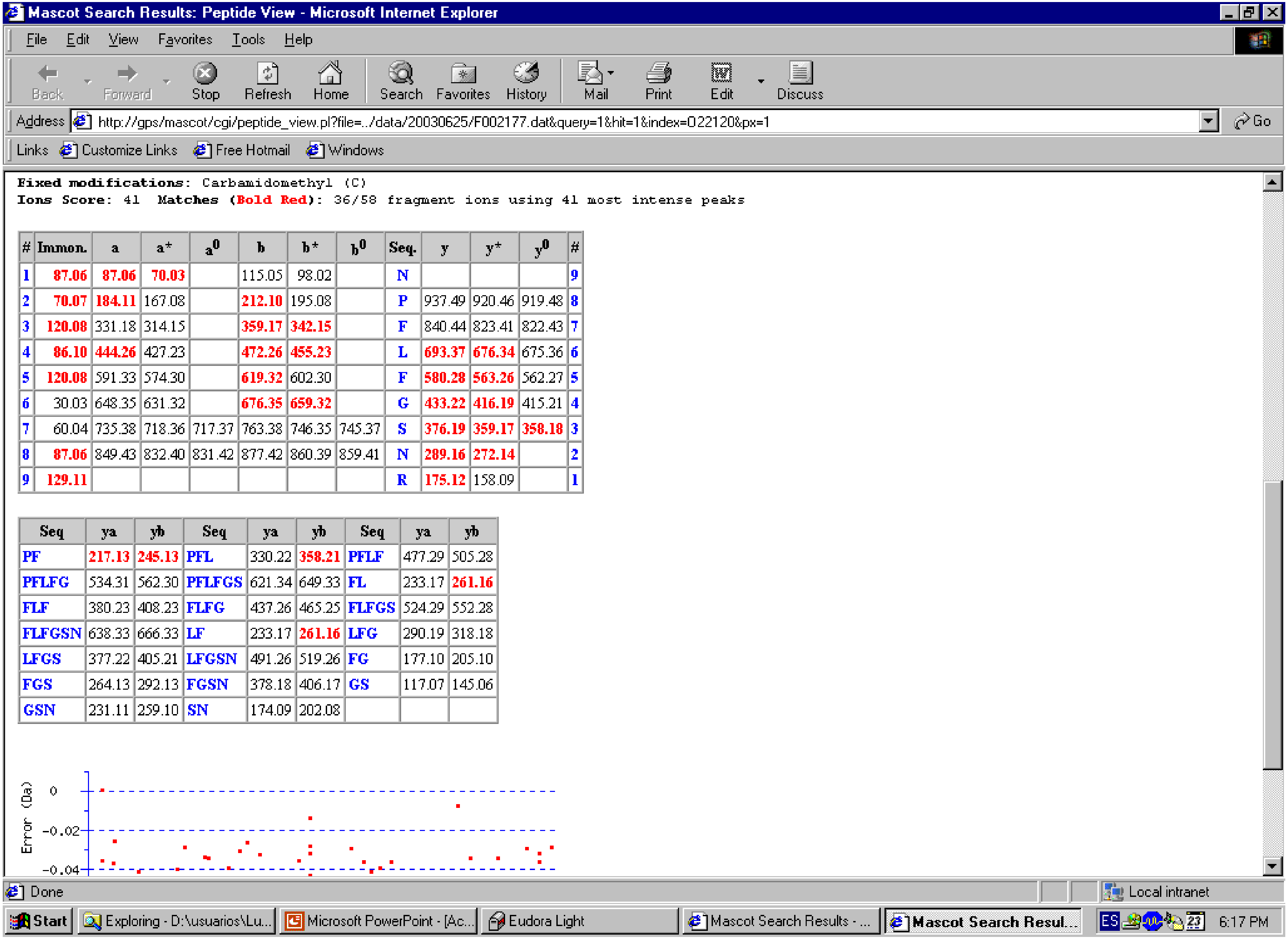
Peptide View

MS/MS Fragmentation of **NPFLFGSNR**
Found in **O22120**, ALPHA SUBUNIT OF BETA CONGLYCININ (FRAGMENT).- Glycine max (Soybean).

Match to Query 1 (1051.49,1+)

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 50 to 950 Da

Monoisotopic mass of neutral peptide (Mr): 1050.52



http://gps/mascot/cgi/peptide_view.pl?file=../data/20030625/F002177.dat&query=1&hit=1&index=022120&px=1

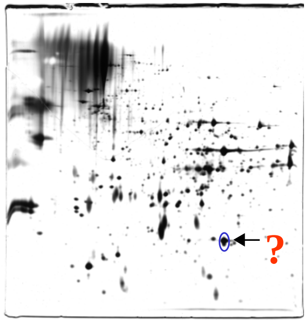
Fixed modifications: Carbamidomethyl (C)

Ions Score: 41 **Matches (Bold Red):** 36/58 fragment ions using 41 most intense peaks

#	Immon.	a	a*	a^0	b	b*	b^0	Seq.	y	y*	y^0	#
1	**87.06**	**87.06**	**70.03**		115.05	98.02		N				9
2	**70.07**	**184.11**	167.08		**212.10**	195.08		P	937.49	920.46	919.48	8
3	**120.08**	331.18	314.15		**359.17**	**342.15**		F	840.44	823.41	822.43	7
4	**86.10**	**444.26**	427.23		**472.26**	**455.23**		L	**693.37**	**676.34**	675.36	6
5	**120.08**	591.33	574.30		**619.32**	602.30		F	**580.28**	**563.26**	562.27	5
6	30.03	648.35	631.32		**676.35**	**659.32**		G	**433.22**	**416.19**	415.21	4
7	60.04	735.38	718.36	717.37	763.38	746.35	745.37	S	**376.19**	**359.17**	**358.18**	3
8	**87.06**	849.43	832.40	831.42	877.42	860.39	859.41	N	**289.16**	**272.14**		2
9	**129.11**							R	**175.12**	158.09		1

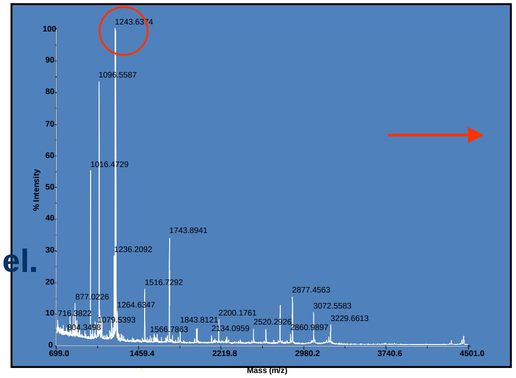
Seq	ya	yb	Seq	ya	yb	Seq	ya	yb
PF	**217.13**	**245.13**	PFL	330.22	**358.21**	PFLF	477.29	505.28
PFLFG	534.31	562.30	PFLFGS	621.34	649.33	FL	233.17	**261.16**
FLF	380.23	408.23	FLFG	437.26	465.25	FLFGS	524.29	552.28
FLFGSN	638.33	666.33	LF	233.17	**261.16**	LFG	290.19	318.18
LFGS	377.22	405.21	LFGSN	491.26	519.26	FG	177.10	205.10
FGS	264.13	292.13	FGSN	378.18	406.17	GS	117.07	145.06
GSN	231.11	259.10	SN	174.09	202.08			

IDENTIFICACIÓN DE PROTEÍNAS MEDIANTE ESPECTROMETRÍA DE MASAS

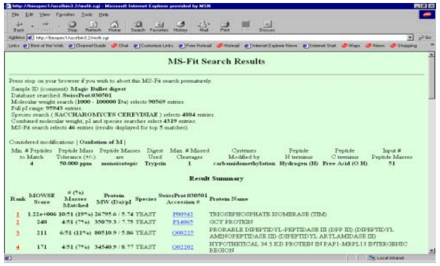


Separación y excisión de la proteina

Digestión con tripsina en gel. Extracción de péptidos.



Obtención de la huella peptídica por espectrometría de masas MALDI-TOF



MS-Fit Search Results

Press stop on your browser if you wish to abort this MS-Fit search prematurely.

Sample ID (comment): **Magic Bullet digest**
Database searched: **SwissProt.030501**
Molecular weight search (**1000 - 100000** Da) selects **90569** entries.
Full pI range: **95943** entries.
Species search (**SACCHAROMYCES CEREVISIAE**) selects **4884** entries.
Combined molecular weight, pI and species searches select **4319** entries.
MS-Fit search selects **46** entries (results displayed for top **5** matches).

Considered modifications: | **Oxidation of M** |

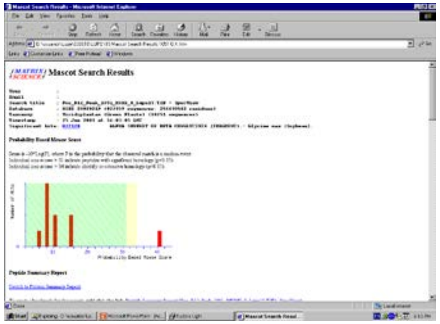
Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest Used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	**50.000 ppm**	**monoisotopic**	**Trypsin**	**1**	**carbamidomethylation**	**Hydrogen (H)**	**Free Acid (O H)**	**51**

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	SwissProt.030501 Accession #	Protein Name
1	1.22e+006	10/51 (19%)	26795.6 / 5.74	YEAST	P00942	TRIOSEPHOSPHATE ISOMERASE (TIM)
2	240	4/51 (7%)	35079.3 / 7.75	YEAST	P14065	GCY PROTEIN
3	211	6/51 (11%)	80510.9 / 5.86	YEAST	Q08225	PROBABLE DIPEPTIDYL-PEPTIDASE III (DPP III) (DIPEPTIDYL AMINOPEPTIDASE III) (DIPEPTIDYL ARYLAMIDASE III)
4	171	4/51 (7%)	34540.9 / 8.77	YEAST	Q02202	HYPOTHETICAL 34.5 KD PROTEIN IN PAP1-MRPL13 INTERGENIC REGION

Búsqueda en bases de datos

Espectrometría de masas en tandem

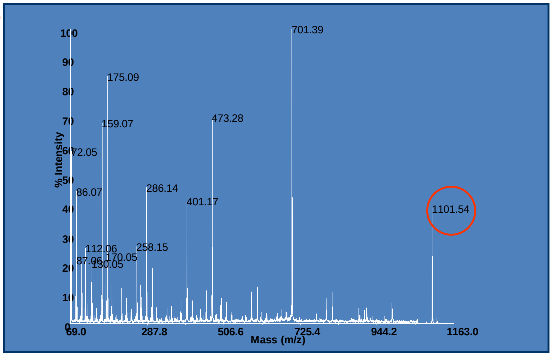


Mascot Search Results

Búsqueda en bases de datos

MPGDVEGLR

Secuenciación *de novo*



Fragmentación del ión peptídico precursor

Ventajas de espectrometría de masas en *tandem* respecto a huella peptídica.

La información de la secuencia es más discriminatoria
que la masa molecular.

Un único péptido de una determinada longitud es suficiente para identificar
una proteína de un genoma secuenciado.

Secuencia *de novo*, se pueden identificar proteínas homólogas de otros
organismos.