

Biological Mass spectrometry in Protein Chemistry

Tuula Nyman
Institute of Biotechnology
tuula.nyman@helsinki.fi

MASS SPECTROMETRY

is an analytical technique that identifies the chemical composition of a compound or sample based on the mass-to-charge ratio of charged particles

ION SOURCE: molecules of interest are ionized



MASS ANALYZER:
ions are separated according to their m/z -ratios



DETECTOR: separated ions are detected



MS of small molecules

-In 1918, Arthur Jeffrey Dempster developed the first modern mass spectrometer, and established the basic theory and design of mass spectrometers that is still used to this day

-1919 Francis Aston constructs the first velocity focusing mass spectrograph with mass resolving power of 130 (1922 Nobel Prize in chemistry)

-The use of a mass spectrometer as the detector in gas chromatography was developed during the 1950s by Roland Gohlke and Fred McLafferty

-Ionization modes: chemical ionization (CI) and electron ionization (EI), not suitable for labile biomolecules



Biological mass spectrometry

-two modes of ionization:
MALDI (matrix assisted laser desorption ionization) and ESI (electrospray ionization)

-developed in 1980's by Michael Karas & Franz Hillenkamp, Koichi Tanaka, John Fenn & Matthias Mann, Peter Roespstorff...

*Nobel-price in Chemistry 2002 to Tanaka and Fenn
'For their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules'*

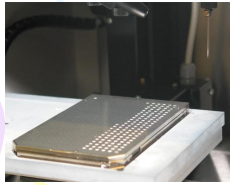
MALDI = matrix assisted laser desorption ionization

TARGET

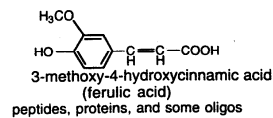
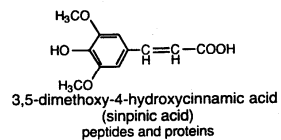
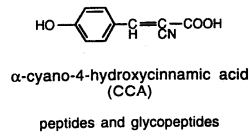
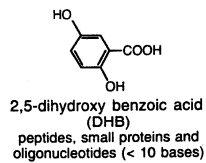
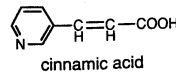
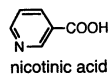
+20 kV

PULSED
LASER

The analyte substance is embedded in a crystallized matrix, which is irradiated by a laser. The power of the laser beam is usually adjusted in a way that it has enough energy to ionize the biomolecules and matrix molecules but does not split the large analyte molecule.



A 384 position MALDI-TOF sample target plate.



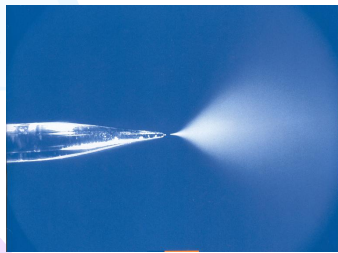
peptides

proteins

Commonly used matrixes for MALDI-TOF mass spectrometry

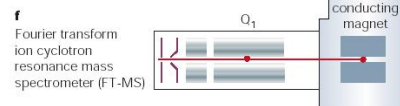
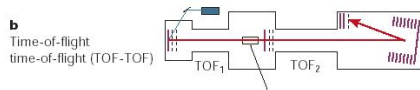
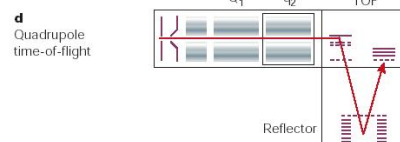
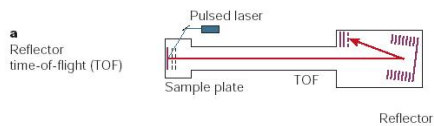
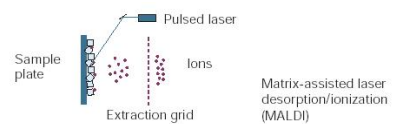
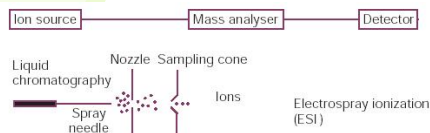
Electrospray ionization (ESI):

Creation of ions by spraying a solution into an electrical field. This process, which belongs to the "soft" ionization techniques, enables the analysis of intact biomolecules, such as e.g. proteins and peptides by mass spectrometry.



Electrospray of peptides in 0.1% FA /ACN from a 15 μm I.D. fused silica glass needle. The liquid flow is 200 nl/min and the needle has a potential of 2000V as compared to the cone inlet of the mass spectrometer.

Mass analyser types



All mass analyzers work in high vacuum

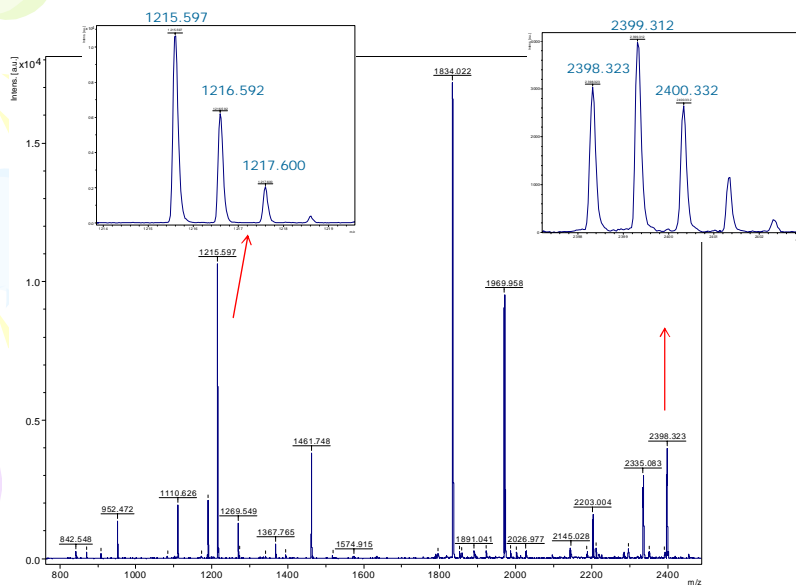
A tandem mass spectrometer is a mass spectrometer that has more than one analyser, in practice usually two.

Tandem mass spectrometry (MS/MS) is used to produce structural information about a compound by fragmenting specific sample ions inside the mass spectrometer and identifying the resulting fragment ions

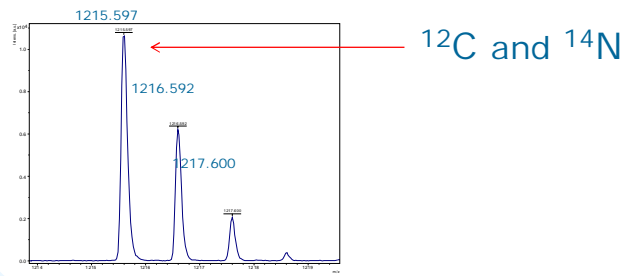
Common MS-instrument types:

- MALDI-TOF and MALDI-TOF/TOF
- ESI-triple quadrupole
- ESI-ion trap
- ESI-hybrid quadrupole TOF
- ESI-ion trap
- ESI-orbitrap

MALDI TOF spectrum of a peptide mixture

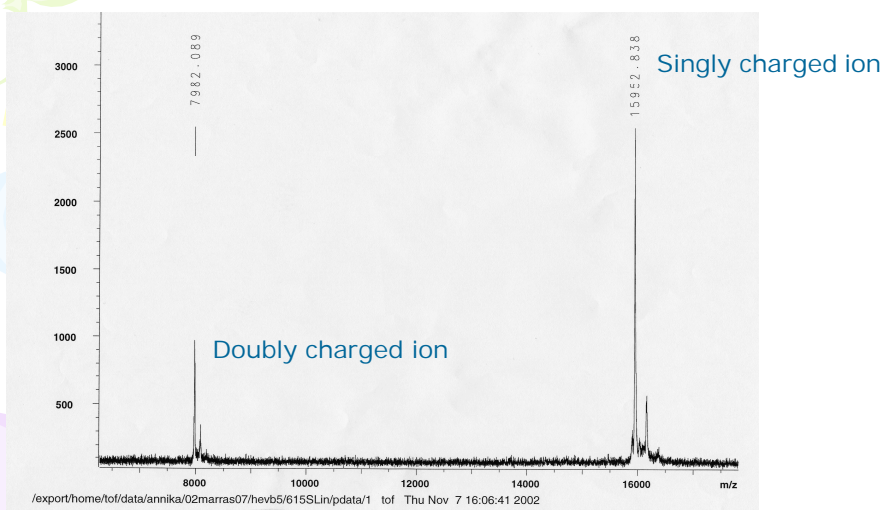


Monoisotopic mass is calculated using the mass of the most abundant natural isotope of each constituent element

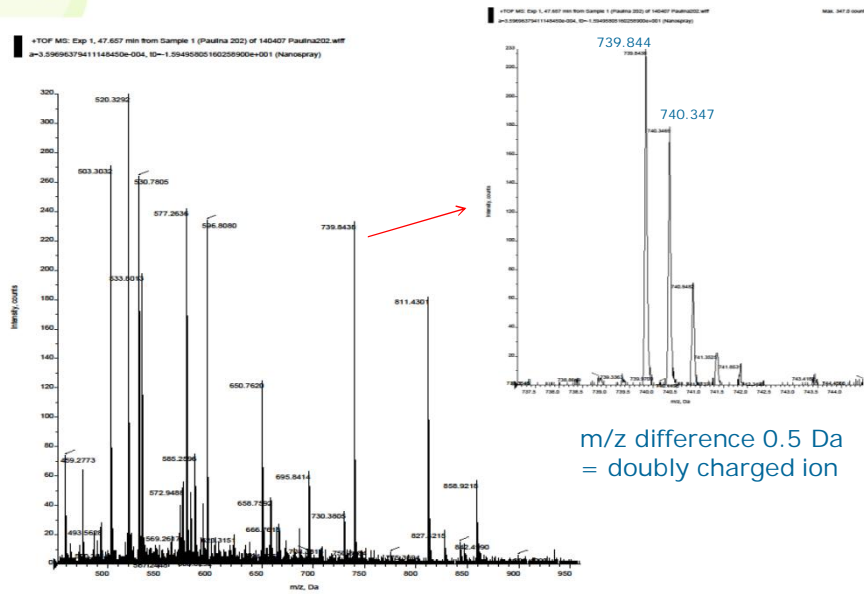


the mass difference between the isotopic peaks is 1 amu (1 Da) as isotopes differ in mass by the addition of 1 neutron which weighs 1 amu

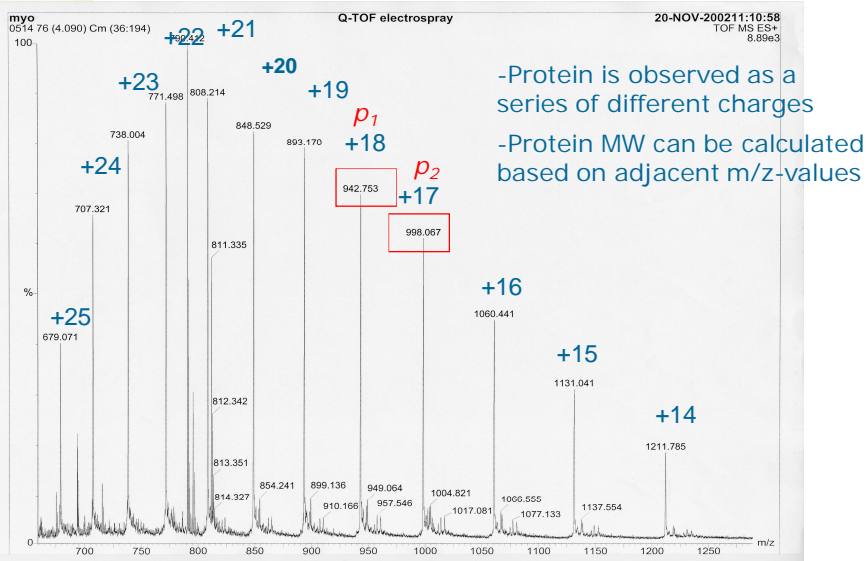
MALDI-TOF spectrum of a protein

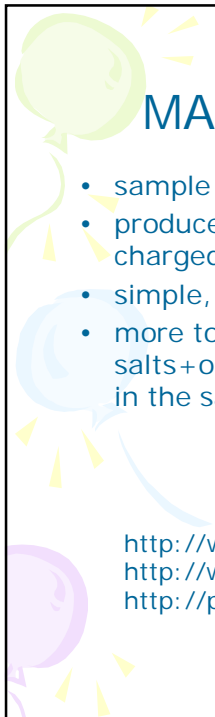


ESI MS spectrum of a peptide mixture

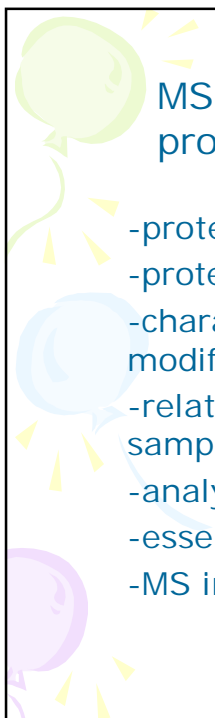


ESI MS spectrum of a protein





MALDI	ESI
<ul style="list-style-type: none">• sample is crystallized• produces mainly singly charged ions• simple, easy-to-use• more tolerant to salts+other contaminants in the sample than ESI	<ul style="list-style-type: none">• liquid sample• produces multiply charged ions• easy to couple with HPLC
<p>http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm http://www.ionsource.com/ http://planetorbitrap.com/</p>	



**MS in protein chemistry/
proteomics/structural biology**

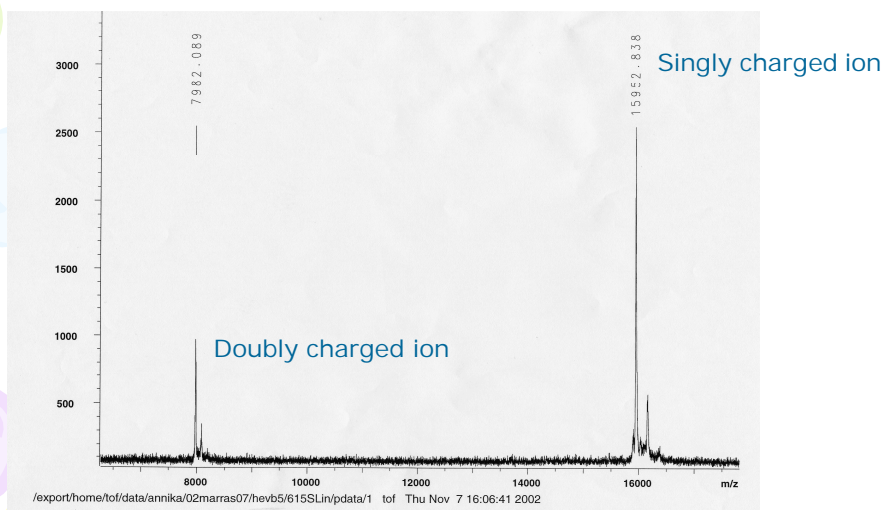
- protein identification
- protein MW determination
- characterisation of post-translational modifications
- relative quantification of proteins between samples
- analysis of protein complexes
- essential role in proteomics
- MS imaging

Protein MW determination by MS (NOT= identification)

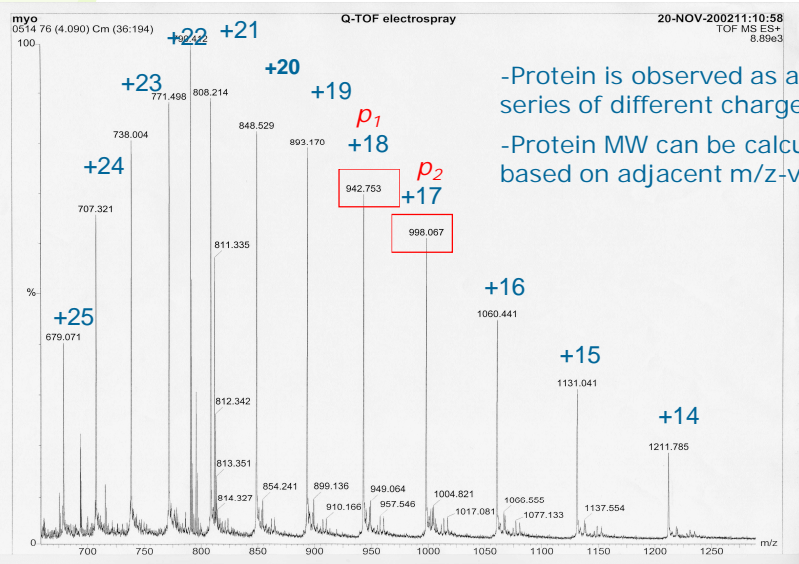
-for MW determination the protein needs to be in solution without salts and detergents

-usually proteins are first purified with RP chromatography before MW measurement

-MALDI TOF MS, linear mode
-accuracy is not as good with ESI MS



Protein MW determination, ESI MS



- Protein is observed as a series of different charges
- Protein MW can be calculated based on adjacent m/z-values

Protein MW calculation from ESI spectra

$$p = m/z$$

$$p_1 = (M_r + z_1) / z_1$$

$$p_2 = [M_r + (z_1 - 1)] / (z_1 - 1)$$

p = a peak in the mass spectrum

m = total mass of an ion

z = total charge

M_r = average mass of the protein

$$p = m/z$$

$$p_1 = (M_r + z_1) / z_1$$

$$p_2 = [M_r + (z_1 - 1)] / (z_1 - 1)$$

If $p_1 = 942.753$ and $p_2 = 998.067$

$$942.753 = (M_r + z_1) / z_1$$

$$942.753 z_1 = M_r + z_1$$

$$941.753 z_1 = M_r$$

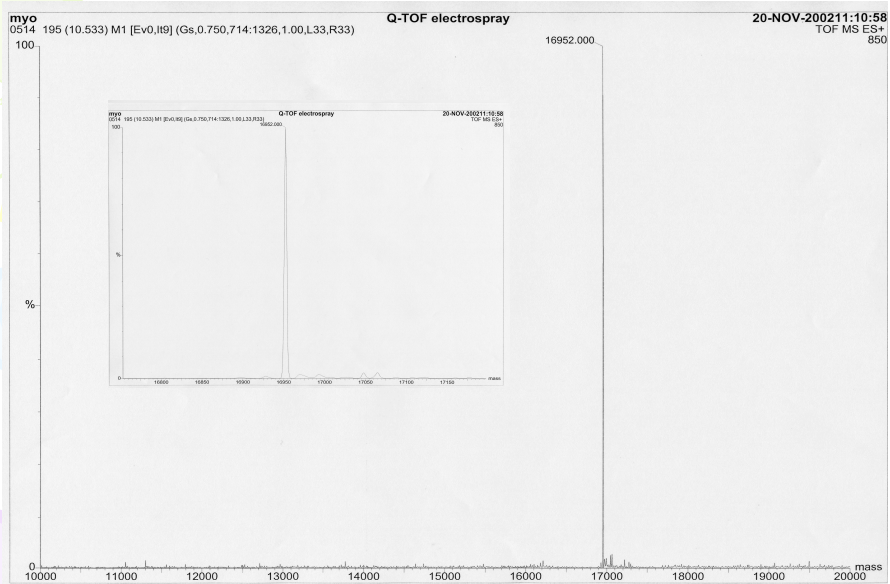
$$998.067 = (941.753 z_1 + z_1 - 1) / (z_1 - 1)$$

$$998.067 z_1 - 997.067 = 942.753 z_1 - 1$$

$$(998.067 - 942.753) z_1 = 996.067$$

$$55.314 z_1 = 996.067 \Rightarrow z_1 = 18.0075$$

$$M_r = 941.753 z_1 = 16\,951.6 \text{ Da}$$



Deconvoluted electrospray mass spectrum of myoglobin



Protein identification by mass spectrometry

- protein of interest is cleaved into peptides with a specific enzyme
- peptides are analyzed by MS (and MS/MS) followed by a database search with the acquired MS (and MS/MS) data

Protein identification methods:

- Peptide mass fingerprinting (PMF)
- Identification based on MS/MS data from one or more peptides



Peptide mass fingerprint (PMF)

A mass spectrum of the peptide mixture resulting from the digestion of a protein by an enzyme, usually measured by MALDI-TOF

Identification based on peptide MW information only

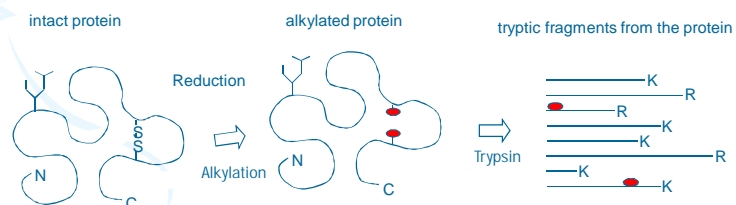
Database search engines create theoretical PMFs for all the proteins in the database and compare these to the measured PMF from protein X

Protein identification based on MS/MS data from one or more peptides

Database search engines create theoretical peptide fragmentation patterns for all the proteins in the database and compare these to the measured MS/MS data

Protein identification by MS

- Protein has to be digested into peptides
- Disulphide bridges need to be reduced and alkylated before digestion





Protein digestion into peptides before MS analysis

Digestion can be done *in-solution* or *in-gel*

The enzyme has to be as specific as possible

Trypsin is most commonly used enzyme:

- very specific, quite cheap
- cleaves peptide bond after lysines and arginines
- tryptic peptides are 'good' for MS analysis because they end up with basic amino acid
- works for both *in-solution* and *in-gel* digestion



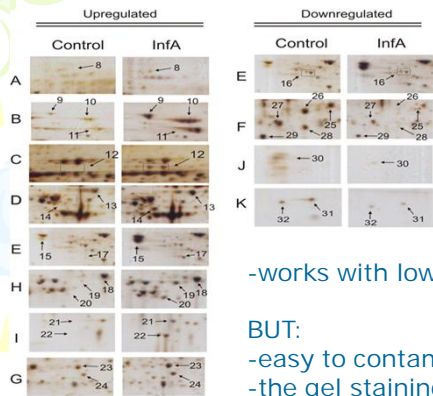
Other enzymes:

- LysC, cleaves after lysines
- LysN, cleaves before lysines
- AspN, cleaves before aspartic acid residues
- V8 protease, cleaves peptide bonds exclusively on the carbonyl side of aspartate and glutamate residues
- possible to do double-digestions

Chemical cleavage:

Cyanogen bromide, cuts after methionines

In-gel digestion



-works with low-femtomolar amounts of proteins

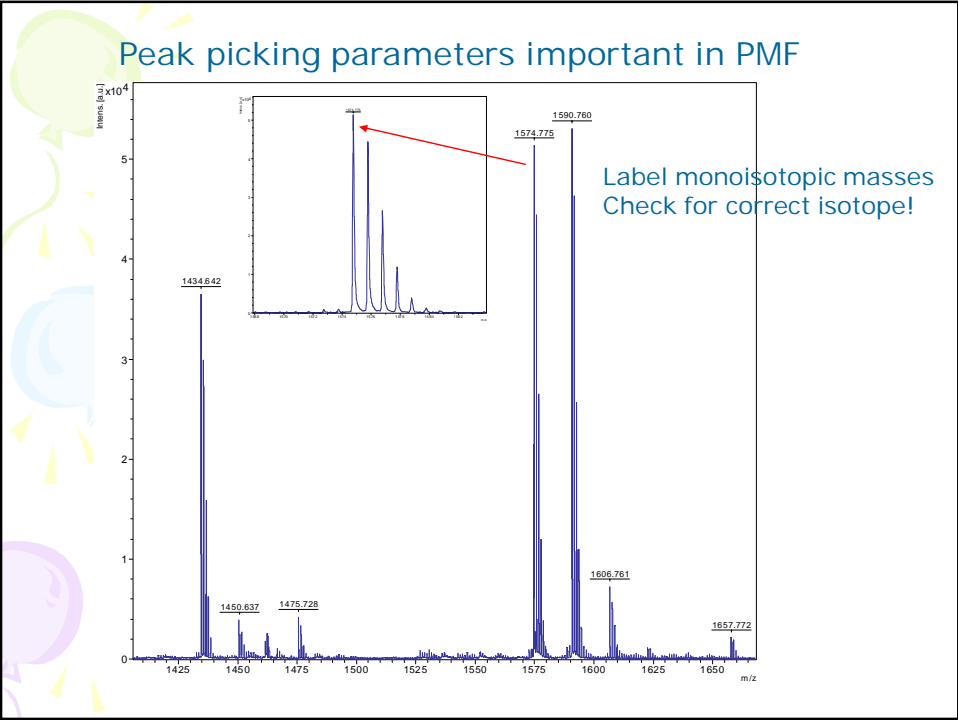
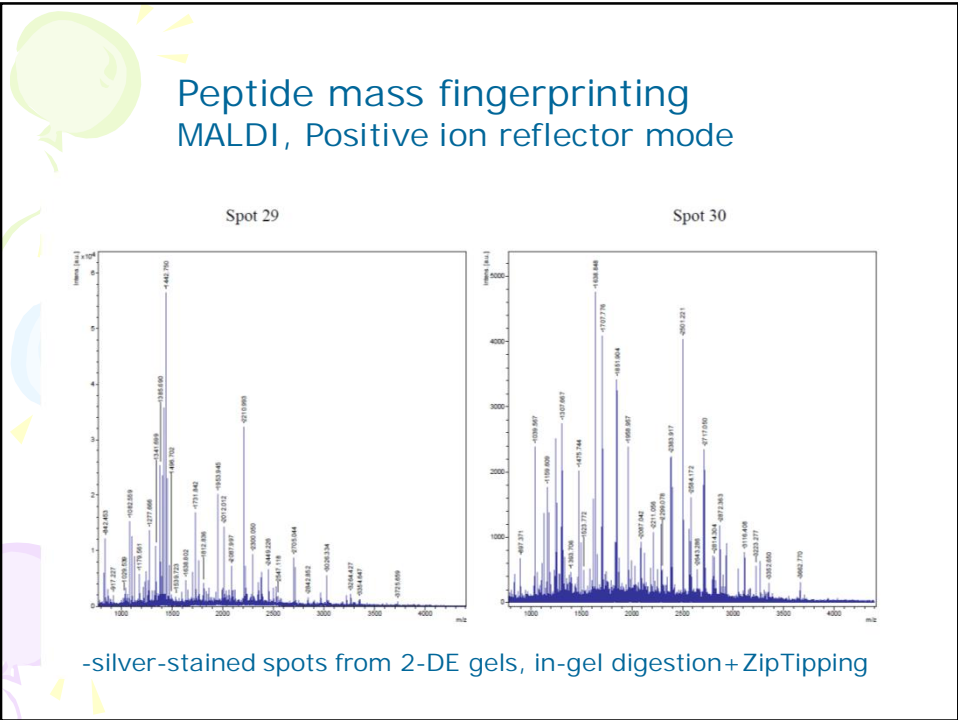
BUT:

- easy to contaminate samples with keratin
- the gel staining protocol needs to be compatible with MS
- silver-staining: fixing with glutaraldehyde cross-links protein into gel matrix after which identification is not possible

Peptide mass fingerprinting

- usually peptides need to be desalted and concentrated before MALDI analysis
- > ZipTips, peptide elution directly onto MALDI target plate
- MALDI matrix included in the elution solution





Publicly available search engines for PMF

- Mascot
- ProteinProspector/ MS-Fit
- PROWL/ ProFound
- Aldente

Databases

Database	Comment
EST	EST divisions of Genbank, (currently EST_human, EST_mouse, EST_others)
MSDB	Comprehensive, non-identical protein database
NCBIInr	Comprehensive, non-identical protein database
SwissProt	High quality, curated protein database

PMF/ database search

MASCOT Peptide Mass Fingerprint

Your name	Tuula	Email	tuula.nyman@helsinki.fi
Search title	dig 33		
Database	NCBIInr		
Taxonomy Homo sapiens (human)		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	<ul style="list-style-type: none"> Biotin (K) Biotin (N-term) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term) 	Variable modifications	<ul style="list-style-type: none"> Oxidation (HW) Oxidation (M) Phospho (ST) Phospho (Y) Propionamide (C)
Protein mass	kDa	Peptide tol. ±	50 ppm
Mass values	<input checked="" type="radio"/> MH ⁺ <input type="radio"/> M _r <input type="radio"/> M-H ⁺	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Data file	842.482722918522 870.515149550309 1085.62503611115 1174.53690038973 1265.55776926686 1393.75672200038		
Decoy	<input type="checkbox"/>	Report top	AUTO hits
Start Search ...		Reset Form	

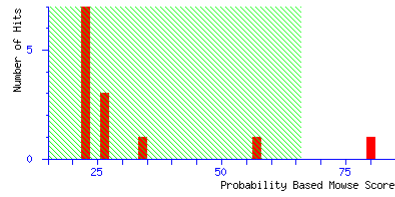
www.matrixscience.com

Mascot Search Results

User : Tuula
 Email : tuula.nyman@helsinki.fi
 Search title : dig 33
 Database : NCBI nr 20090912 (9680073 sequences: 3307708198 residues)
 Taxonomy : Homo sapiens (human) (223942 sequences)
 Timestamp : 17 Sep 2009 at 07:28:13 GMT
 Top Score : 80 for [gi|109096484](#), PREDICTED: tubulin, alpha, ubiquitous isoform 10 [Macaca mulatta]

Probability Based Mowse Score

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Protein scores greater than 66 are significant ($p < 0.05$).



Concise Protein Summary Report

Format As: [Help](#)
 Significance threshold $p <$ Max. number of hits

- [gi|109096484](#) Mass: 46797 Score: 80 Expect: 0.0022 Queries matched: 12
 PREDICTED: tubulin, alpha, ubiquitous isoform 10 [Macaca mulatta]
[gi|193786502](#) Mass: 46725 Score: 80 Expect: 0.0022 Queries matched: 12
 unnamed protein product [Homo sapiens]
[gi|193787715](#) Mass: 46825 Score: 80 Expect: 0.0022 Queries matched: 12
 unnamed protein product [Homo sapiens]
[gi|34740335](#) Mass: 50804 Score: 76 Expect: 0.0052 Queries matched: 12
 tubulin, alpha 1B [Mus musculus]
[gi|73996547](#) Mass: 46781 Score: 69 Expect: 0.029 Queries matched: 11
 PREDICTED: similar to tubulin, alpha 1 isoform 9 [Canis familiaris]
[gi|109096516](#) Mass: 37707 Score: 66 Expect: 0.054 Queries matched: 10
 PREDICTED: alpha tubulin isoform 2 [Macaca mulatta]
[gi|158259731](#) Mass: 50788 Score: 66 Expect: 0.058 Queries matched: 11
 unnamed protein product [Homo sapiens]

(MATRIX)
(SCIENCE) **Mascot Search Results**

Protein View

Match to: [gi|193786502](#) Score: 80 Expect: 0.0022
 unnamed protein product [Homo sapiens]

Nominal mass (M_1): 46725; Calculated pI value: 4.99
 NCBI BLAST search of [gi|193786502](#) against nr
 Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Homo sapiens](#)

Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Number of mass values searched: 24
 Number of mass values matched: 12
 Sequence Coverage: 40%

Matched peptides shown in **Bold Red**

```

1  MPSDKTIGGG DSENFHFESE TGAGKHVPRA VFVDLEPTVI DEVRTGTYFQ
51  LFHPEQLITG KEDAANNYAR GHYTIQKEII DLVLDRIKLI ADQCTGLQGF
101 LVFHSFGGGT GSGFTSLIME RLSVDYGGKS KLEFSIYPAP QVSTAVVEPY
151 NSILTTHTTL EHSDFCFMVD NEAIVDICRR NLDIERPTTY NLRRLISQIV
201 SSITASLRFD GALNVDLTFE QTNLWPPYRI HFPLATYAPV ISAEKAYHEQ
251 LSVAEITHAC FEPANQWVKC DPRHGKMAC CLLXRGDVPV KDVNAIATI
301 KTKRSIQFVD WCPYTGKVGCI NYQPTTVVPG GDLAKVQRAN CHLSNTTATA
351 EAWARLDHGF DLMYAKRAFV HUVVGGEMEE GEFSEAREDM AALEKDYEEV
401 GVDSVEGEGG EGEVEY
  
```

Show predicted peptides also

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Usually not many missed cleavages

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss Sequence
6 - 25	2007.8886	2006.8813	2006.8858	-2	0 K.TIGGGDSENFHFESETGAGK.H
30 - 44	1701.9195	1700.9122	1700.8985	8	0 R.AVFVDLEPTVIDEVR.T
50 - 61	1410.7704	1409.7631	1409.7667	-3	0 R.QLFHPEQLITGK.E
50 - 70	2415.1967	2414.1894	2414.1978	-3	1 R.QLFHPEQLITGKEDAANNYAR.G
78 - 86	1085.6250	1084.6178	1084.6128	5	0 K.EIIDLVLDL.R
181 - 194	1718.8900	1717.8827	1717.8747	5	0 R.NLDIERPTYTHLNR.L
230 - 245	1756.9589	1755.9517	1755.9559	-2	0 R.DHFPLATYAPVISAER.A
246 - 269	2766.2690	2765.2617	2765.2789	-6	0 K.AYHEQLSVAEITHACFEPANQWVK.C Oxidation (M)
277 - 285	1265.5578	1264.5505	1264.5403	8	0 K.YMACLLYR.G Oxidation (M)
305 - 317	1584.7560	1583.7487	1583.7443	3	0 R.SIQFVDWCPYTGK.V
318 - 335	1824.9830	1823.9757	1823.9782	-1	0 K.VGINYQPTTVVPGDLAK.V
356 - 366	1396.6967	1395.6894	1395.6857	3	1 R.LDHKFDLMYAK.R Oxidation (M)

No match to: 842.4827, 870.5151, 1174.5368, 1393.7567, 1616.7599, 1715.9463, 1778.9576, 1846.9822, 2029.8757, 2211

Mass accuracy critical

-check for known contaminants (these should be removed before search)
 -possibility to do 2nd pass search

'Normal' search, max 1 missed cleavage allowed

Match to: **TM2_HUMAN** Score: 61 Expect: 0.43
Tropomyosin beta chain OS=Homo sapiens GN=TM2 PE=1 SV=1

Nominal mass (M₀): 32945; Calculated pI value: 4.66

NCBI BLAST search of **TM2_HUMAN** against nr
Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Homo sapiens](#)

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 41
Number of mass values matched: 13
Sequence Coverage: 33%

Matched peptides shown in **Bold Red**

```

1 MDAIKYKQK IMLDVENAID RAQQAADKQ QAEDRCRQLK ERQQAQKQL
51 KGTEDVEKY SESVFAEKK LEQAERQTD AADAVASLNR RIQLVESELD
101 RAQRALATAL QRLEASAAK DESERGNVI ENRANDEEK MELQMSLKE
151 ARIARSDSR KYEVARSLV ILRQLERSE ERAPVAEKK GDLEELKIV
201 TINDLSLAQ ADGVYDSEK YEEIIMLKE KLRAATRAE FAERSVAKLE
251 RYTDIDLEQV YAGQSYVAI SEELDINALD IISL

```

Max 4 missed cleavages allowed

Protein View

Match to: **TM2_BOVIN** Score: 120 Expect: 5e-07
Tropomyosin beta chain OS=Bos taurus GN=TM2 PE=2 SV=1

Nominal mass (M₀): 32931; Calculated pI value: 4.66

NCBI BLAST search of **TM2_BOVIN** against nr
Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Bos taurus](#)

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 41
Number of mass values matched: 31
Sequence Coverage: 75%

Matched peptides shown in **Bold Red**

```

1 MDAIKYKQK IMLDVENAID RAQQAADKQ QAEDRCRQLK ERQQAQKQL
51 KGTEDVEKY SESVFAEKK LEQAERQTD AADAVASLNR RIQLVESELD
101 RAQRALATAL QRLEASAAK DESERGNVI ENRANDEEK MELQMSLKE
151 ARIARSDSR KYEVARSLV ILRQLERSE ERAPVAEKK GDLEELKIV
201 TINDLSLAQ ADGVYDSEK YEEIIMLKE KLRAATRAE FAERSVAKLE
251 RYTDIDLEQV YAGQSYVAI SEELDINALD IISL

```

Start	End	Observed	Mr(expt)	Nr(calcd)	ppm	Miss	Sequence
13	20	2043.7676	2044.7527	2043.6222	-129	0	R.LIDRSDILRBAQVAADK.Q
36	48	1631.5830	1630.5727	1630.7984	-137	1	R.CQLRERQQAQK.R
38	48	1343.4100	1344.4627	1342.6729	-142	0	R.CQLRERQQAQK.R
49	55	1355.6660	1354.5987	1354.6381	-29	2	R.KLRLRDEYK.V Phospho (ST)
52	76	2868.9540	2867.5967	2868.2512	-127	3	R.GYEDRPVRESVVDVQSLRQAK.R
60	70	1442.6070	1442.5927	1442.2187	44	1	R.YEYSDVAEKK.L Phospho (ST); Pho
77	90	1460.5320	1459.5247	1459.7267	-138	1	R.RATDAADAVASLNR.R
77	91	1616.6070	1615.5927	1615.8276	-141	2	R.RATDAADAVASLNR.I
78	90	1332.4600	1331.4527	1331.6317	-134	0	R.ATDAADAVASLNR.R
78	91	1488.5360	1487.5287	1487.7128	-137	1	R.ATDAADAVASLNR.I
91	101	1339.5710	1338.5637	1338.7467	-131	1	R.IQLVESELD.R
91	105	1883.7500	1882.7427	1882.5860	-129	2	R.IQLVESELDRAQER.L
92	101	1243.4870	1242.4827	1242.6454	-128	0	R.IQLVESELD.R
92	105	1727.6700	1726.6627	1726.8845	-129	1	R.IQLVESELDRAQER.L
106	125	2201.8610	2200.8537	2201.1172	-120	2	R.LATAIQRLERARADRSR.G
113	128	1898.7320	1897.7127	1897.7021	-40	2	R.LERASAALRESDR.V Oxidation (M)
124	149	1980.8380	1979.8287	1979.3208	-120	2	R.ARIARSDSRKYEVARSLV.R
124	149	1998.8880	1995.8727	1995.3210	-128	2	R.ARIARSDSRKYEVARSLV.R Oxidation (M)
127	151	2310.1290	2309.1217	2309.4028	-99	4	R.ISESDRQLRQLRQALASDNRK.V RFD
128	157	1917.6270	1916.6127	1916.8702	-138	2	R.RIARSDSRKYEVARSLV.R
148	148	1098.9630	1097.2827	1097.7402	-127	1	R.KLVLLRSELK.S
188	192	1739.7490	1738.7427	1738.9788	-132	2	R.KLVLLRSELK.S
189	148	3170.2090	3169.2227	3169.8534	-122	0	R.KLVLLRSELK.S
189	189	2282.9620	2284.1827	2282.2081	-122	2	R.KLVLLRSELK.S
206	226	2263.8890	2262.8927	2263.1380	-126	3	R.SLQAPAKNTYKDKYDESLK.L
210	223	2319.1260	2313.1487	2313.2488	-127	4	R.SLQAPAKNTYKDKYDESLK.L
214	221	2214.0290	2213.2217	2213.4024	-3	3	R.YTYGAKYDESLKLEK.L Phospho (S)
218	221	1794.6600	1793.6527	1793.8334	-134	2	R.DNVESELDLEK.L
218	221	1874.7030	1873.6987	1873.8237	-99	2	R.DNVESELDLEK.L Phospho (Y)
218	228	2622.0290	2621.2217	2621.2420	-123	4	R.DNVESELDLEKLRGAKTR.A
222	244	3238.9210	3237.2127	3237.1180	-121	0	R.TIDRDLRQVAGK.H

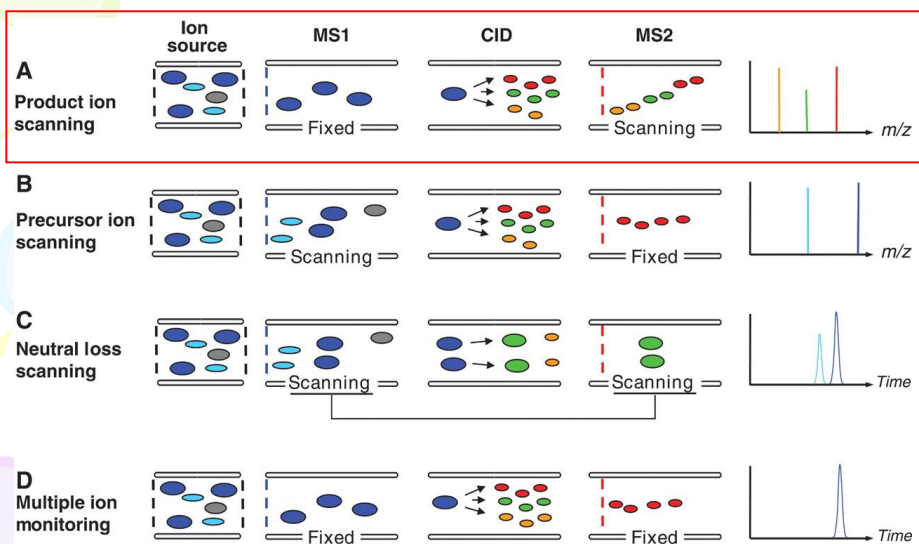
PEPTIDE MASS FINGERPRINTING

- 'quick and easy'
- requires
 - a very specific enzyme
 - optimized digestion+ desalting protocols
 - internal/close external calibration of MALDI spectra
- works only for proteins which are already in the databases as protein sequences
- not suitable for complex protein mixtures

Protein identification based on MS/MS data from one or more peptides

- MALDI-TOF/TOF or nanoLC-ESI-MS/MS analysis
- suitable for (complex) protein mixtures, especially when combined with LC separation of peptides before MS/MS

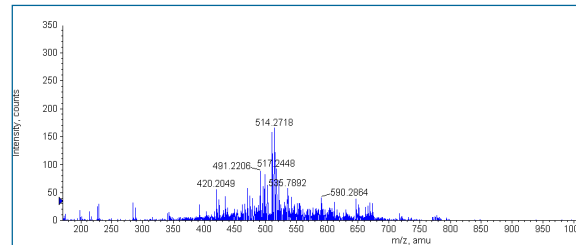
Tandem mass spectrometry scan types



Product ion scanning

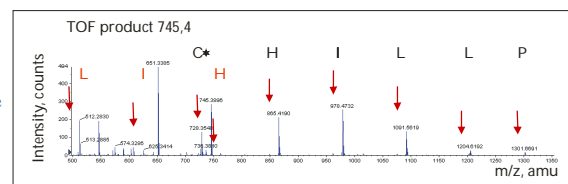
MS scan:

The ions are first separated according to their m/z ratios

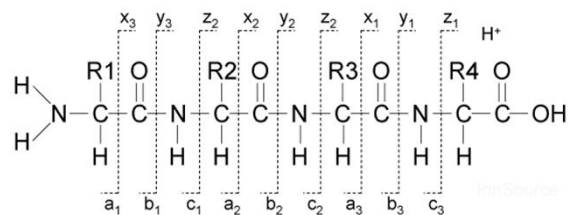


MS/MS scan:

Peptides with certain m/z -ratio are selected and fragmented inside the mass analyzer, and the m/z -ratios of the fragment ions are measured



Peptide Fragmentation Nomenclature



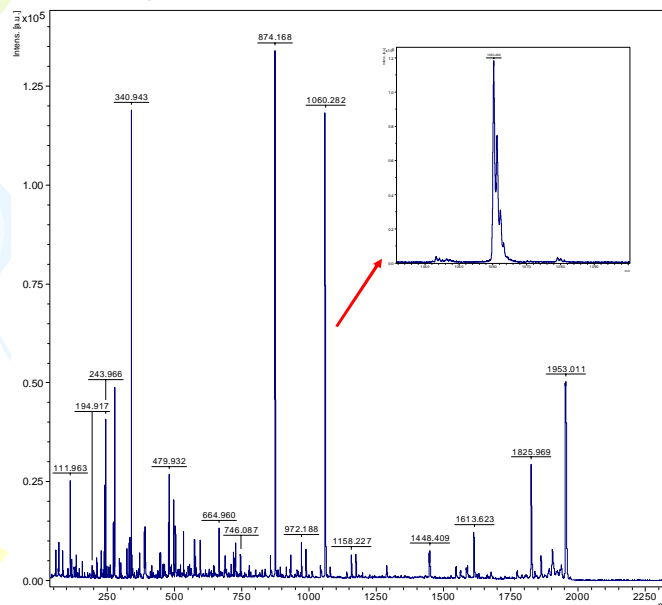
Peptides do not fragment sequentially, the fragmentation events are somewhat random.

The most common peptide fragments observed in low energy collisions are a, b and y ions. The b ions appear to extend from the amino terminus (N-terminus), and y ions appear to extend from the carboxyl terminus (C-terminus).

Protein identification with MALDI-TOF/TOF

- first, PMF with MALDI-TOF
- next, selected precursor ions can be fragmented (TOF/TOF analysis)
- MALDI produces singly charged parent ions
 - product ion spectra not as easy to interpret as in ESI-MS/MS
- database search with both PMF and MS/MS information

MALDI-TOF/TOF fragment ion spectra from parent ion m/z 1952



Sequence Query: One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc.

Mascot Search Results

User: dm111a
 Result: dm111a_rosschotelhelinski.f1
 Search title: 08_421
 MS data file: DATA.TXT
 Database: SwissProt 57.7 (49729 sequences): 175274722 residues
 Timestamp: 18 Sep 2009 at 06:30:43 EDT
 Protein hits: SECI_YEAST Protein transport protein SECI OS=Saccharomyces cerevisiae GN=SECI PE=1 SV=1

Probability Based Mouse Score

Ion score is $-10 \log(P)$, where P is the probability that the observed match is a random event. Individual ion scores > 39 indicate identity or extensive homology ($p < 0.05$). Protein scores are derived from ion scores as a non-probabilistic basis for ranking protein hits.

Peptide Summary Report

Significance threshold: 0.05
 Standard scoring: MadPIT scoring
 Show pop-ups: Suppress pop-ups
 Select All: Search Selected

1. SECI_YEAST Mass: 8714 Score: 70 Queries matched: 1
 Protein transport protein SECI OS=Saccharomyces cerevisiae GN=SECI PE=1 SV=1
 Check to include this hit in error tolerant search

Query	Observed	Mr (exp1)	Mr (calc)	ppm	Miss	Score	Repeat	Rank	Peptide
1	1952.8700	1951.8627	1951.9284	-33.66	0	81	3.9e-06	1	YVINCIDAFMVRPPK

Search Parameters

Mascot Search Results: Peptide View

Peptide View

MS/MS Fragmentation of YVINCIDAFMVRPPK
 Found in SECI_YEAST, Protein transport protein SECI OS=Saccharomyces cerevisiae GN=SECI PE=1 SV=1
 Match to Query 1: 1951.862724 from(1952.870000,1-)
 Data file: DATA.TXT

Click mouse within plot area to zoom in by factor of two about that point
 Or, Plot from 1000 to 1900 Da Full range

Monoisotopic mass of neutral peptide Mr (calc): 1951.9284
 Fixed modifications: Carbamidomethyl (C)
 Ion Score: 81 Expect: 3.9e-06
 Matches (Bold face): 33/259 fragment ions using 30 most intense peaks

#	Inmon.	a	a+	a0	b	b+	b0	d	Seq.	v	w	w'	y	y'	y0	#
1	136.0757	136.0757			164.0706				Y							16
2	87.0553	250.1186	233.0921		278.1135	261.0870			N	1730.8353	1729.8400		1789.8724	1772.8458	1771.8618	15
3	86.0964	363.2027	346.1761		391.1976	374.1710			I	1617.7512	1630.7716	1644.7873	1675.8295	1658.8029	1657.8189	14
4	87.0553	477.2456	460.2191		505.2405	488.2140			N	1503.7083	1502.7130		1562.7454	1545.7188	1544.7348	13
5	123.0430	637.2763	620.2497		665.2713	648.2446			C	1343.6776	1342.6824		1448.7025	1431.6759	1430.6919	12
6	86.0964	750.3603	733.3338		776.3552	761.3287			I	1230.5936	1243.6140	1257.6296	1288.6718	1271.6451	1270.6611	11
7	88.0393	865.3873	848.3607		847.3767	833.3522	876.3556	875.3716	D	1115.5666	1114.5714		1175.5878	1158.5612	1157.5772	10



Protein identification using nanoLC-MS/MS

- 50-75 μm i.d. RP-columns, 200 nl/min
→ no need to split the effluent before MS

- DDA= data dependent analysis, can be fully automated

- suitable for complex protein mixtures,
possibility to identify hundreds of proteins in one run



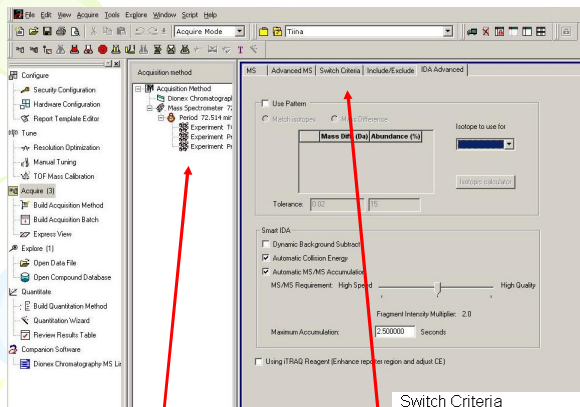
DDA = Data Dependent Acquisition
(IDA = Information Dependent Acquisition)

- fully automated experiment, first MS scan followed by two or more product ion scans

- the acquisition software is set to choose certain types of ions for fragmentation and to use 'suitable' collision energy for this ion

- in ESI tryptic peptides have usually 2-4 charges

DDA experiment

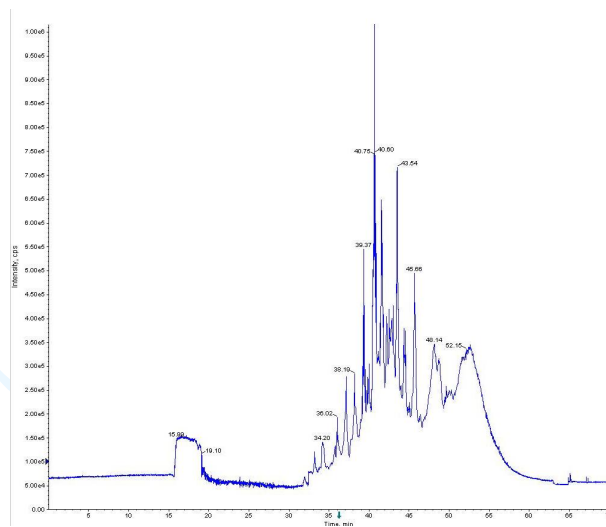


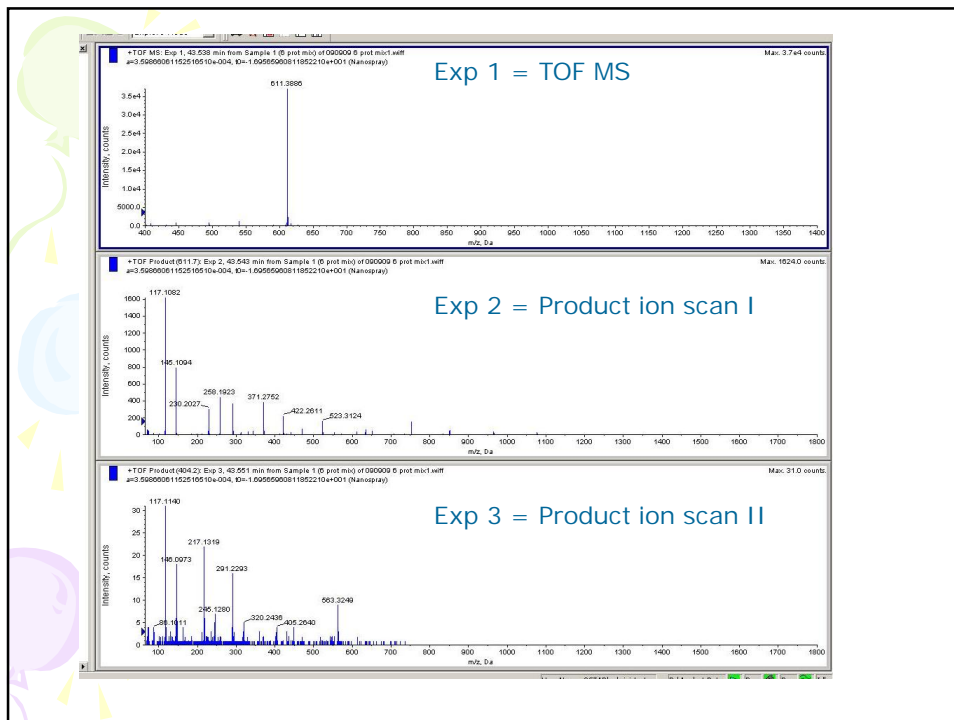
Switch Criteria
 For ions greater than: 400.000 m/z
 For ions smaller than: 1400.000 m/z
 With charge state: 2 to 4
 Which exceeds:
 Exclude former target ions:
 Switch after:
 Mass Defect Filter: No
 Ions Tolerance:

10 counts
 For: 60 seconds
 1 spectra
 50.000 mDa

- Exp 1 = TOF MS
- Exp 2 = Product ion scan from parent I
- Exp 3 = Product ion scan from parent II

TIC = total ion current





Search engines for (LC-)MS/MS data

- Mascot, Sequest, OMSSA etc
- the programs take the fragment ion spectrum of a peptide as input and score it against theoretical fragmentation patterns constructed for peptides from the searched database.
- in practise the user is often limited to use those search engines which accept the data format from the mass spec used
- mzXML is a open data format for storage and exchange of mass spec data
- raw, proprietary file formats from most vendors can be converted to the open mzXML format

Mascot Search Results

Dear :
 Email :
 Search title :
 Database : SwissProt 57.7 (497293 sequences; 175274722 residues)
 Timestamp : 11 Sep 2009 at 07:10:14 GMT
 Enzyme : Trypsin
 Fixed modifications : Carbamidomethyl (C)
 Variable modifications : Oxidation (M), Phospho (ST), Phospho (Y)
 Mass values : Monoisotopic
 Protein mass : Unrestricted
 Peptide mass tolerance : ± 50 ppm
 Fragment mass tolerance : ± 0.2 Da
 Max missed cleavages : 1
 Instrument type : ESI-QTOF-TOF
 Number of queries : 1107
 Protein hits : [P48679|LAMA_RAT](#) Lamin-A OS=Rattus norvegicus GM=Lama PE=1 SV=1
[P48679|LAMA_MOUSE](#) Lamin-A/C OS=Mus musculus GM=Lama PE=1 SV=2
[P15598|KCTD7_HUMAN](#) Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GM=KRT2 PE=1 SV=2
[P02341|TYRP1_HUMAN](#) Tyrosinase OS=Homo sapiens GM=TYR1 PE=1 SV=1
[P15598|KCTD7_HUMAN](#) Keratin, type II cytoskeletal 1 OS=Homo sapiens GM=KRT1 PE=1 SV=6
[P15527|KIF1C_HUMAN](#) Keratin, type I cytoskeletal 9 OS=Homo sapiens GM=KIF9 PE=1 SV=3
[P13345|KIF10_HUMAN](#) Keratin, type I cytoskeletal 10 OS=Homo sapiens GM=KIF10 PE=1 SV=5
[P13347|KIF10_HUMAN](#) Keratin, type II cytoskeletal 5 OS=Homo sapiens GM=KIF5 PE=1 SV=3
[Q0B511|GMP13_FOSM1](#) Stress-70 protein, mitochondrial OS=Homo sapiens GM=HSP70 PE=2 SV=1
[P18888|CPT2_HUMAN](#) Carnitine O-palmitoyltransferase 2, mitochondrial OS=Rattus norvegicus GM=Cpt2 PE=1 SV=1
[Q04948|TYRO_TROV1](#) Tyrosinase beta chain OS=Homo sapiens GM=TYR2 PE=2 SV=1
[P15240|HSP90_HUMAN](#) Hsp90 class B OS=Homo sapiens GM=HSP90 PE=1 SV=1
[Q0R273|HSP90_HUMAN](#) Hsp90 class A OS=Homo sapiens GM=HSP90 PE=1 SV=2
[P02331|KIF14_HUMAN](#) Keratin, type I cytoskeletal 14 OS=Homo sapiens GM=KIF14 PE=1 SV=4
[P02709|ALDH1_HUMAN](#) Aldehyde dehydrogenase 1 OS=Homo sapiens GM=ALDH1 PE=1 SV=4
[P15287|FANCG_HUMAN](#) Fancin C OS=Homo sapiens GM=FANCG PE=1 SV=1
[P15165|HDC_HUMAN](#) Dermcidin OS=Homo sapiens GM=HDC PE=1 SV=2
[P15169|HDC_MOUSE](#) Protein S100-A8 OS=Homo sapiens GM=S100A8 PE=1 SV=1

Select Summary Report

Format As: Select Summary (protein hits) [Help](#)

Significance threshold p: Max. number of hits:

Standard scoring MascIT scoring Ions score or expect cut-off Show sub-sets

Show pop-ups Suppress pop-ups Sort unassigned: Require bold red

Mascot Search Results

Protein View

Match to: [P48679|LAMA_RAT](#) Score: 889
 Lamin-A OS=Rattus norvegicus GM=Lama PE=1 SV=1

Nominal mass (M_n): 74564; Calculated pI value: 6.54
 NCBI BLAST search of [P48679|LAMA_RAT](#) against nr
 Unformatted sequence listing for pasting into other applications

Taxonomy: [Rattus norvegicus](#)

Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Sequence Coverage: 47%

Matched peptides shown in **Bold Red**

```

1 METPQSRFF RSQAQSTPT LAFPTVRIQ EKKDLGKLD ELAVIVR
51 SMTYDRIAS LATESEVY RSEVYDIAA YEAQDGAAS TLGVQVQAA
101 RLQLELSEV EPFPLKARN TVRSDLLAA QAPLQLEAL LSRREALST
151 ALBERTDSD RLDGLQVRA YKLRADGAK RUGGDEKPT VDSNGLTGI
201 YSELDPDGI YSELLETER RSTVELVID NGQRETER LADALQELA
251 QHEQVQYK RLKLDYTKK LKAPQDQER HNSVYGAAS ELQQRKID
301 ELKAPQDQ RLKAVDAS KLEKELAKR NQTFELAKR YKLRADGAA
351 RPKQLKRTQ RLDGLKALD MSLKVRKLL EDEKELRLD PFTQQRKQ
401 ELKSDSPQ GGVYDTRK LKREKREKTF IGAATYDQV ANVEDEKRE
451 PVLKADQNE DQKSDNLIK RQKEDDPIAT YFPFPTLIK AGQVTKAS
501 QGAKREPTT DLNVAQYK GCTTFLKAL DAKSREYAK RKLVDLSTP
551 ENKSDSDQ DLKLRDQAS RQKSDKQAS PMSSTPLQK QYQDPAKDA
601 ARDSDAQDQ SIKSDQDAS VYVTRFRYV GSGGDFDQ MIVTRVILG
651 NSSTVQVQK KCIK
  
```

Show predicted peptides also

Sort Peptides By: Residue Number Increasing Mass Decreasing Mass

Start	End	Observed	Mr (exact)	Mr (calc)	ppm	Miss Sequence
12	25	690.3506	3358.6867	3358.6790	6	R. RSQAQSTPLSTPT .I (Ions score 20)
12	25	720.3309	1438.6472	1438.6453	1	R. RSQAQSTPLSTPT .I Phospho (ST) (Ions score 33)
29	41	943.9397	1428.7973	1428.8005	-2	R. LQKEDGQELDGR .L (Ions score 14)
29	41	866.8704	1130.5263	1130.5304	3	R. RLQLELSEV .S (Ions score 45)
42	48	425.2464	848.4782	848.4795	3	R. LAVTIDR .V (Ions score 41)
51	60	545.2760	1088.5263	1088.5462	-10	R. RLSTENALR .L (Ions score 21)
51	60	545.2802	1088.5408	1088.5462	-0	R. RLSTENALR .L (Ions score 21)
51	60	585.2447	1148.5149	1148.5125	2	R. RLSTENALR .L Phospho (ST) (Ions score 33)
63	72	574.7959	1147.5732	1147.5721	4	R. RTESEVVER .E (Ions score 20)
63	72	574.7966	1147.5786	1147.5721	6	R. RTESEVVER .E (Ions score 1)
63	72	614.7697	1227.5248	1227.5384	-11	R. RTESEVVER .E Phospho (ST) (Ions score 20)
63	72	614.7715	1227.5285	1227.5384	-8	R. RTESEVVER .E Phospho (ST) (Ions score 18)
79	90	431.8848	828.6325	828.6364	-3	R. RLQLELSEV .S (Ions score 45)
102	108	415.7532	828.4918	828.4909	1	R. RLQLELSEV .S (Ions score 33)
102	108	415.7543	828.4940	828.4909	4	R. RLQLELSEV .S (Ions score 14)
120	133	586.3246	1170.6385	1170.6397	2	R. RRLKLANQK .R (Ions score 35)
134	144	622.3481	1242.7216	1242.7183	3	R. RLQLELSEV .E (Ions score 45)
145	155	560.3099	1118.6053	1118.5813	21	R. RKALSTALR .R (Ions score 15)
145	156	638.3548	1274.6951	1274.6830	9	R. RKALSTALR .R (Ions score 14)
145	156	678.3318	1364.6490	1364.6490	-0	R. RKALSTALR .S Phospho (ST) (Ions score 4)
155	164	591.8111	1181.6777	1181.6640	3	R. RKALSTALR .S (Ions score 14)
172	180	451.2539	900.4932	900.4914	2	R. RKALSTALR .K (Ions score 45)
172	181	515.2941	1028.5716	1028.5864	-13	R. RKALSTALR .Q (Ions score 14)
181	189	580.8085	1159.6024	1159.6019	0	R. RKQGDGR .R (Ions score 21)

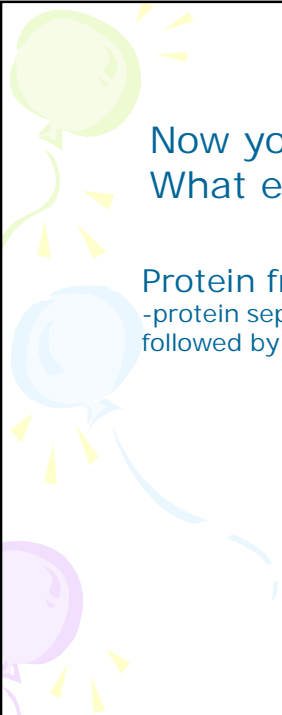
Ion scores from individual MS/MS spectra





Protein identification from complex mixtures using nanoLC-MS/MS

- produces huge amounts of raw data
- requires efficient data processing tools
- different database search programs can produce different results from the same raw data
- false discovery rate estimation
 - peptide level
 - protein level



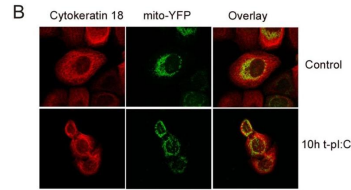
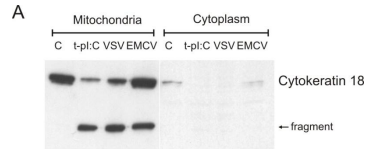
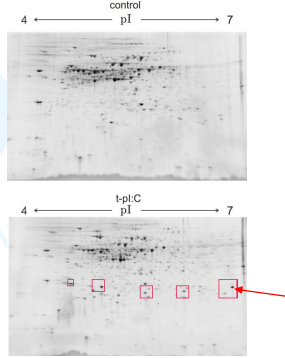
Now you're protein is identified What else can we find in the data?

Protein fragmentation

- protein separation by SDS-PAGE/2-DE followed by in-gel digestion and MS analysis

Cytokeratin-18 is cleaved during viral infection, and fragments localize onto mitochondria

Mitochondrial proteomes



Mascot Search Results. Protein View

(MATRIX) Mascot Search Results

Protein View

Search to: F05783|K1C18_HUMAN Score: 927
 Keratin, type I cytokeletal 18 (Homo sapiens GN=KRT18 PE=1 SV=2)
 Nominal mass (kDa): 48029; Calculated pI value: 5.34
 NCBI BLAST search of [F05783|K1C18_HUMAN](#) against nr
 Unformatted [sequence listing](#) for pasting into other applications
 Taxonomy: [Homo sapiens](#)
 Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Sequence Coverage: 40%

Matched peptides shown in **Red**

```

1 MSFTIRSTPS ENYRSLGSVQ APTSGAPFVS SASVYVAGG GSGRSLRVER
51 SIFTFGQMS GGLATVAGG LAGRGGIQWE KETKTMQSLNDR LASTLRVER
101 LLETNRRLS KIRBLEKGG FQVRMSHYF KIIEELPAQI FANTVDNARI
151 VLIQNNRRLA ADFRKYKET ELAMRQSVEN DIBGLRKYD DYNITRLQLE
201 TRIRALRREL LRFQDREHE WGLQQAQAS SOLVVEYAP RQGLALDIA
251 DIRAQVDELA RNSREELRKY NSQIEESTT VVTQSAEVS AERTITELR
301 RTVQSLRIDL DGRNRFASL ENSLREYAR VALQMEQMG ILHLSEELA
351 QFAAGQGRFA QVVALRMYR VGLREIARY RRLRLEGGDF VLGGALDSSN
401 SMTICRITIT FRVVDGRVVS ETNDEKVLK
    
```

Show predicted peptides also

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calcd)	Delta	Miss	Sequence
7 - 14	488.2270	974.4394	974.4458	-0.0063	0	R.STFSTNYR.S (ions_score_53)
7 - 14	488.2307	974.4468	974.4458	0.0010	0	R.STFSTNYR.S (ions_score_30)
7 - 14	488.2311	974.4476	974.4458	0.0018	0	R.STFSTNYR.S (ions_score_3)
7 - 14	488.2316	974.4486	974.4458	0.0028	0	R.STFSTNYR.S (ions_score_13)
7 - 14	488.2481	974.4816	974.4458	0.0358	0	R.STFSTNYR.S (ions_score_11)
15 - 45	952.1442	2853.4109	2853.4005	0.0103	0	R.SLQVQASVQARFVSSAAVDTGGGGGSR.L (ions_score_86)
56 - 81	765.0267	2292.0581	2292.0838	-0.0257	0	R.GGMSSGLATVAGGLAVMGGIQNEK.E 2 Oxidation (M) (ions_score_13E)
56 - 81	765.0359	2292.0858	2292.0838	0.0020	0	R.GGMSSGLATVAGGLAVMGGIQNEK.E 2 Oxidation (M) (ions_score_15)
56 - 90	846.6363	3382.5279	3382.5552	-0.0272	1	R.GGMSSGLATVAGGLAVMGGIQNEKTMQSLNDR.L 3 Oxidation (M) (ions_score_1)
56 - 90	1128.5232	3382.5479	3382.5552	-0.0073	1	R.GGMSSGLATVAGGLAVMGGIQNEKTMQSLNDR.L 3 Oxidation (M) (ions_score_1)
91 - 97	419.1995	836.3845	836.4392	-0.0547	0	R.LASVYLR.V (ions_score_5)
91 - 97	419.2254	836.4362	836.4392	-0.0030	0	R.LASVYLR.V (ions_score_43)
91 - 97	419.2257	836.4368	836.4392	-0.0024	0	R.LASVYLR.V (ions_score_53)
91 - 97	419.2261	836.4377	836.4392	-0.0015	0	R.LASVYLR.V (ions_score_13)
100 - 107	502.7557	1003.4969	1003.5046	-0.0077	1	R.SLETENRR.L (ions_score_1)

Based on 2-DE and MS data:
N-terminal fragment of the identified protein

Post-translational modifications

TAP-tag purification of protein complexes,
on-beads digestion with double enzyme,
first LysC 2h, then trypsin o/n

Mascot Search Results

User: [redacted]
 Search title: [redacted]
 Database: SwissProt 54.3 (189338 sequences; 104773129 residues)
 Timestamp: 8 Sep 2008 at 07:54:34 GMT
 Enzyme: Trypsin
 Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
 Mass values: Unrestricted
 Protein Mass: Unrestricted
 Peptide Mass Tolerance: ± 50 ppm
 Fragment Mass Tolerance: ± 0.2 Da
 Max Missed Cleavages: 2
 Instrument type: MS1-QMAP-TOF
 Number of queries: 1057
 Protein hits:

- KIN2_YEAST** Serine/threonine-protein kinase KIN2 - *Saccharomyces cerevisiae* (Baker's yeast)
- KIC1_HUMAN Keratin, type II cytoskeletal 1 - *Homo sapiens* (Human)
- KIC2_HUMAN Keratin, type II cytoskeletal 2 - *Homo sapiens* (Human)
- KIC3_HUMAN Keratin, type II cytoskeletal 3 - *Homo sapiens* (Human)
- KIC4_HUMAN Keratin, type II cytoskeletal 4 - *Homo sapiens* (Human)
- KIC5_HUMAN Keratin, type II cytoskeletal 5 - *Homo sapiens* (Human)
- KIC6_HUMAN Keratin, type II cytoskeletal 6 - *Homo sapiens* (Human)
- KIC7_HUMAN Keratin, type II cytoskeletal 7 - *Homo sapiens* (Human)
- KIC8_HUMAN Keratin, type II cytoskeletal 8 - *Homo sapiens* (Human)
- KIC9_HUMAN Keratin, type II cytoskeletal 9 - *Homo sapiens* (Human)
- KIC10_HUMAN Keratin, type II cytoskeletal 10 - *Homo sapiens* (Human)
- KIC11_HUMAN Keratin, type II cytoskeletal 11 - *Homo sapiens* (Human)
- KIC12_HUMAN Keratin, type II cytoskeletal 12 - *Homo sapiens* (Human)
- KIC13_HUMAN Keratin, type II cytoskeletal 13 - *Homo sapiens* (Human)
- KIC14_HUMAN Keratin, type II cytoskeletal 14 - *Homo sapiens* (Human)
- KIC15_HUMAN Keratin, type II cytoskeletal 15 - *Homo sapiens* (Human)
- KIC16_HUMAN Keratin, type II cytoskeletal 16 - *Homo sapiens* (Human)
- KIC17_HUMAN Keratin, type II cytoskeletal 17 - *Homo sapiens* (Human)
- KIC18_HUMAN Keratin, type II cytoskeletal 18 - *Homo sapiens* (Human)
- KIC19_HUMAN Keratin, type II cytoskeletal 19 - *Homo sapiens* (Human)
- KIC20_HUMAN Keratin, type II cytoskeletal 20 - *Homo sapiens* (Human)
- KIC21_HUMAN Keratin, type II cytoskeletal 21 - *Homo sapiens* (Human)
- KIC22_HUMAN Keratin, type II cytoskeletal 22 - *Homo sapiens* (Human)
- KIC23_HUMAN Keratin, type II cytoskeletal 23 - *Homo sapiens* (Human)
- KIC24_HUMAN Keratin, type II cytoskeletal 24 - *Homo sapiens* (Human)
- KIC25_HUMAN Keratin, type II cytoskeletal 25 - *Homo sapiens* (Human)
- KIC26_HUMAN Keratin, type II cytoskeletal 26 - *Homo sapiens* (Human)
- KIC27_HUMAN Keratin, type II cytoskeletal 27 - *Homo sapiens* (Human)
- KIC28_HUMAN Keratin, type II cytoskeletal 28 - *Homo sapiens* (Human)
- KIC29_HUMAN Keratin, type II cytoskeletal 29 - *Homo sapiens* (Human)
- KIC30_HUMAN Keratin, type II cytoskeletal 30 - *Homo sapiens* (Human)
- KIC31_HUMAN Keratin, type II cytoskeletal 31 - *Homo sapiens* (Human)
- KIC32_HUMAN Keratin, type II cytoskeletal 32 - *Homo sapiens* (Human)
- KIC33_HUMAN Keratin, type II cytoskeletal 33 - *Homo sapiens* (Human)
- KIC34_HUMAN Keratin, type II cytoskeletal 34 - *Homo sapiens* (Human)
- KIC35_HUMAN Keratin, type II cytoskeletal 35 - *Homo sapiens* (Human)
- KIC36_HUMAN Keratin, type II cytoskeletal 36 - *Homo sapiens* (Human)
- KIC37_HUMAN Keratin, type II cytoskeletal 37 - *Homo sapiens* (Human)
- KIC38_HUMAN Keratin, type II cytoskeletal 38 - *Homo sapiens* (Human)
- KIC39_HUMAN Keratin, type II cytoskeletal 39 - *Homo sapiens* (Human)
- KIC40_HUMAN Keratin, type II cytoskeletal 40 - *Homo sapiens* (Human)
- KIC41_HUMAN Keratin, type II cytoskeletal 41 - *Homo sapiens* (Human)
- KIC42_HUMAN Keratin, type II cytoskeletal 42 - *Homo sapiens* (Human)
- KIC43_HUMAN Keratin, type II cytoskeletal 43 - *Homo sapiens* (Human)
- KIC44_HUMAN Keratin, type II cytoskeletal 44 - *Homo sapiens* (Human)
- KIC45_HUMAN Keratin, type II cytoskeletal 45 - *Homo sapiens* (Human)
- KIC46_HUMAN Keratin, type II cytoskeletal 46 - *Homo sapiens* (Human)
- KIC47_HUMAN Keratin, type II cytoskeletal 47 - *Homo sapiens* (Human)
- KIC48_HUMAN Keratin, type II cytoskeletal 48 - *Homo sapiens* (Human)
- KIC49_HUMAN Keratin, type II cytoskeletal 49 - *Homo sapiens* (Human)
- KIC50_HUMAN Keratin, type II cytoskeletal 50 - *Homo sapiens* (Human)
- KIC51_HUMAN Keratin, type II cytoskeletal 51 - *Homo sapiens* (Human)
- KIC52_HUMAN Keratin, type II cytoskeletal 52 - *Homo sapiens* (Human)
- KIC53_HUMAN Keratin, type II cytoskeletal 53 - *Homo sapiens* (Human)
- KIC54_HUMAN Keratin, type II cytoskeletal 54 - *Homo sapiens* (Human)
- KIC55_HUMAN Keratin, type II cytoskeletal 55 - *Homo sapiens* (Human)
- KIC56_HUMAN Keratin, type II cytoskeletal 56 - *Homo sapiens* (Human)
- KIC57_HUMAN Keratin, type II cytoskeletal 57 - *Homo sapiens* (Human)
- KIC58_HUMAN Keratin, type II cytoskeletal 58 - *Homo sapiens* (Human)
- KIC59_HUMAN Keratin, type II cytoskeletal 59 - *Homo sapiens* (Human)
- KIC60_HUMAN Keratin, type II cytoskeletal 60 - *Homo sapiens* (Human)
- KIC61_HUMAN Keratin, type II cytoskeletal 61 - *Homo sapiens* (Human)
- KIC62_HUMAN Keratin, type II cytoskeletal 62 - *Homo sapiens* (Human)
- KIC63_HUMAN Keratin, type II cytoskeletal 63 - *Homo sapiens* (Human)
- KIC64_HUMAN Keratin, type II cytoskeletal 64 - *Homo sapiens* (Human)
- KIC65_HUMAN Keratin, type II cytoskeletal 65 - *Homo sapiens* (Human)
- KIC66_HUMAN Keratin, type II cytoskeletal 66 - *Homo sapiens* (Human)
- KIC67_HUMAN Keratin, type II cytoskeletal 67 - *Homo sapiens* (Human)
- KIC68_HUMAN Keratin, type II cytoskeletal 68 - *Homo sapiens* (Human)
- KIC69_HUMAN Keratin, type II cytoskeletal 69 - *Homo sapiens* (Human)
- KIC70_HUMAN Keratin, type II cytoskeletal 70 - *Homo sapiens* (Human)
- KIC71_HUMAN Keratin, type II cytoskeletal 71 - *Homo sapiens* (Human)
- KIC72_HUMAN Keratin, type II cytoskeletal 72 - *Homo sapiens* (Human)
- KIC73_HUMAN Keratin, type II cytoskeletal 73 - *Homo sapiens* (Human)
- KIC74_HUMAN Keratin, type II cytoskeletal 74 - *Homo sapiens* (Human)
- KIC75_HUMAN Keratin, type II cytoskeletal 75 - *Homo sapiens* (Human)
- KIC76_HUMAN Keratin, type II cytoskeletal 76 - *Homo sapiens* (Human)
- KIC77_HUMAN Keratin, type II cytoskeletal 77 - *Homo sapiens* (Human)
- KIC78_HUMAN Keratin, type II cytoskeletal 78 - *Homo sapiens* (Human)
- KIC79_HUMAN Keratin, type II cytoskeletal 79 - *Homo sapiens* (Human)
- KIC80_HUMAN Keratin, type II cytoskeletal 80 - *Homo sapiens* (Human)
- KIC81_HUMAN Keratin, type II cytoskeletal 81 - *Homo sapiens* (Human)
- KIC82_HUMAN Keratin, type II cytoskeletal 82 - *Homo sapiens* (Human)
- KIC83_HUMAN Keratin, type II cytoskeletal 83 - *Homo sapiens* (Human)
- KIC84_HUMAN Keratin, type II cytoskeletal 84 - *Homo sapiens* (Human)
- KIC85_HUMAN Keratin, type II cytoskeletal 85 - *Homo sapiens* (Human)
- KIC86_HUMAN Keratin, type II cytoskeletal 86 - *Homo sapiens* (Human)
- KIC87_HUMAN Keratin, type II cytoskeletal 87 - *Homo sapiens* (Human)
- KIC88_HUMAN Keratin, type II cytoskeletal 88 - *Homo sapiens* (Human)
- KIC89_HUMAN Keratin, type II cytoskeletal 89 - *Homo sapiens* (Human)
- KIC90_HUMAN Keratin, type II cytoskeletal 90 - *Homo sapiens* (Human)
- KIC91_HUMAN Keratin, type II cytoskeletal 91 - *Homo sapiens* (Human)
- KIC92_HUMAN Keratin, type II cytoskeletal 92 - *Homo sapiens* (Human)
- KIC93_HUMAN Keratin, type II cytoskeletal 93 - *Homo sapiens* (Human)
- KIC94_HUMAN Keratin, type II cytoskeletal 94 - *Homo sapiens* (Human)
- KIC95_HUMAN Keratin, type II cytoskeletal 95 - *Homo sapiens* (Human)
- KIC96_HUMAN Keratin, type II cytoskeletal 96 - *Homo sapiens* (Human)
- KIC97_HUMAN Keratin, type II cytoskeletal 97 - *Homo sapiens* (Human)
- KIC98_HUMAN Keratin, type II cytoskeletal 98 - *Homo sapiens* (Human)
- KIC99_HUMAN Keratin, type II cytoskeletal 99 - *Homo sapiens* (Human)
- KIC100_HUMAN Keratin, type II cytoskeletal 100 - *Homo sapiens* (Human)

Select Summary Report

Format As: [dropdown] Select Summary (protein hits) [dropdown] Help

Significance threshold: p < [input type="text" value="0.05"] Max. number of hits [input type="text" value="AUTO"]

Standard scoring Mascot scoring Less score or expect cut-off Show sub-sets

Show pop-ups Suppress pop-ups Sort unassigned [Decreasing Score] Require bold red

Mascot Search Results

Protein View

Match to: KIN2_YEAST Score: 791
 Serine/threonine-protein kinase KIN2 - *Saccharomyces cerevisiae* (Baker's yeast)

Nominal mass (M₀): 128601; Calculated pI value: 9.43
 NCBI BLAST search of KIN2_YEAST against nr
 Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Saccharomyces cerevisiae](#)

Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M), Phospho (ST), Phospho (Y)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is F
 Sequence Coverage: 33%

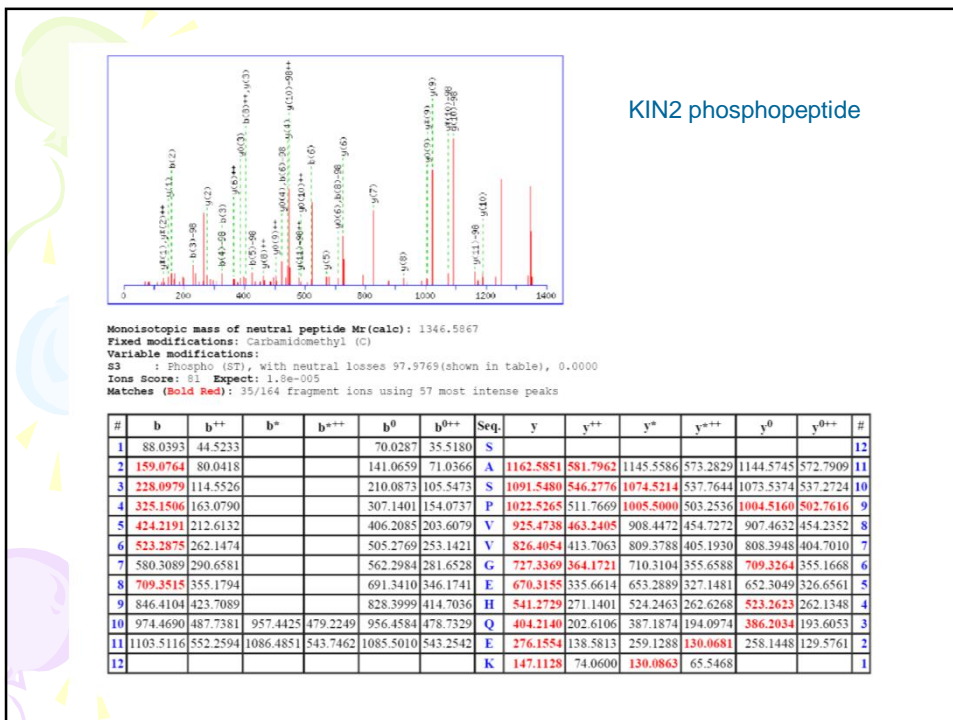
Matched peptides shown in Bold Red

```

1 MNPWVADYL VNNPFTSHG GSLPTEAF NDRVLRAPL IRMGKQSGP
51 RNDQQAPLH PFDIKHGHE QAGRQNDAS RFGAVELRQ FRRSLGDWE
101 FLETVAGSM GWNVLVHRQI THEICVIVK VRAKAYLKH QBSLSPKNE
151 SEILMQRRL EKEIARDKRT VEARSLQIL YPHICRLFE MCTMNRHFM
201 LFEIVSGQL LDYIQNGSL KEHARRKPAR GIASALQLL ANNVHRDLK
251 REIMISRS EKIKIIFGL NEDVPRQLN FPGCVLFA FHLKAPPT
301 GPEVDISFG IYLVLVCGH VFDENSSI LNEKIKGVV DVFMSIEV
351 ISLLTRMIV DLRRLATLN VVEHPWNRG VDFKAPSV NRVELPEMI
401 DSQVLEMYR LEFIDIDP RSLIRLVE VETIQSGEV WDLKNAKL
451 SGLNNVYLS STAQQILQW HTSPFQSG WEDFENED FLAVYPLLS
501 IYHLYEMVA RKLAKLQRQ ALAQAQAGQ RQQQQVALG TKVLMNNSP
551 DINTVGRSPQ NEVVPNGIF QVAIGTGT ENNTNSNKE FLNVPFKL
601 TPEQATSP TGRSSDRT ELNGVLRTP VEVSGEYQQ SASFVGEHQ
651 RNTVGGIFP RISQSGQR FTRQELFE RPTFMNS NEIKIVFNS
701 RSPTSDIIS SARVGVVW NVDVWQKA KWTAPAFIS SVSGRNSDL
751 PALPQNELI VQKQRKLLQ ENLRLQIND NDNVNVAVV DGINDNDSW
801 YLSVFGKRL NPSARAKVVG HARRESLKT RFPFALPF SIMTNDNGL
851 GEAKVLEMYR VSNMSTVPE DDTYSNWN NLETIVYQK LTRQLLEEA
901 SNAPVGRMS IDVFWMLK GFSVQVTS KLFIVRNI ISVTKMSID
951 FKEVKGKFC VQRFSEITA AVFVITIGV GLDGRAMD QNSLDSQLS
1001 SYHTASASG RNSIKRQGS YKQGNHLE TPLANTHRQ NSSIPSPHY
1051 GNQSHGSGE LSHMSLVQV QDDILITER AQMINVWQV TEQNTSPGK
1101 ERFPVREIS IVKRVIGLA GVEPKVSGN FVNLKELASV ILEKLNL
  
```

Cleaved peptides shown in Bold Red

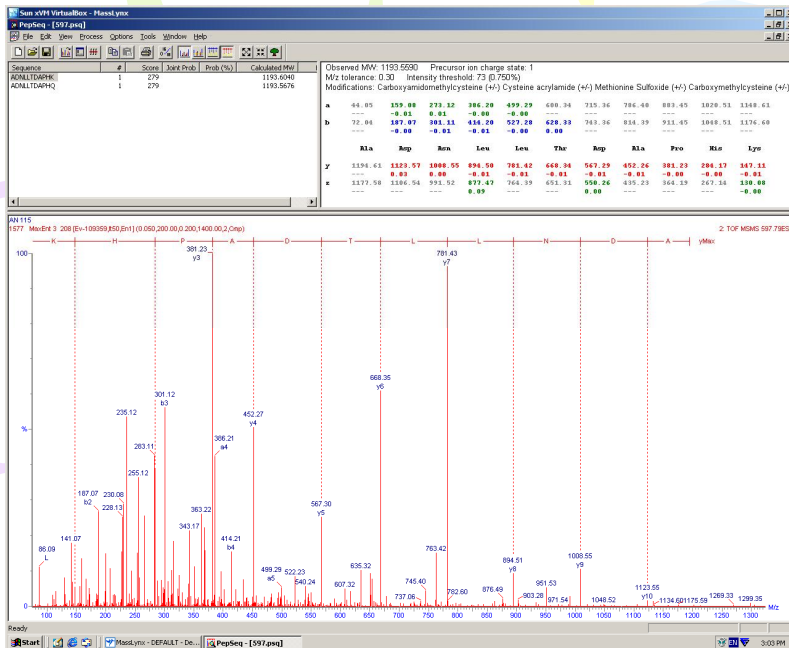
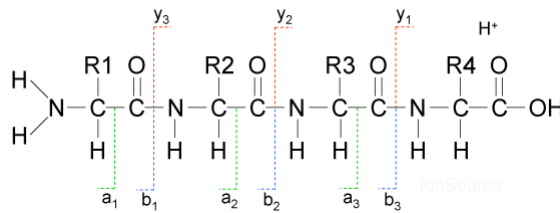
Start - End	Observed	Mr (expt)	Mr (calc)	ppm	Miss Sequence
2 - 16	866.4372	1730.8598	1730.8376	13	0 M.FNPNADYLVWPFNR.T (Ions score 82)
20 - 34	774.9776	1547.7406	1547.7216	12	0 K.GGSLSPFPFAFDTR.V (Ions score 66)
52 - 66	833.4293	1664.8441	1664.8192	15	0 R.MDQQAPIMFPAIR.Q (Ions score 37)
141 - 148	447.2544	892.4942	892.4767	20	0 K.QNSLSPK.N (Ions score 10)
141 - 148	487.2387	972.4629	972.4430	21	0 K.QNSLSPK.N Phospho (ST) (Ions score 5)
149 - 156	495.2514	988.4883	988.4825	6	0 K.NKSEILDR.Q (Ions score 65)
370 - 379	427.8884	1280.6435	1280.6084	27	0 K.NVVEHPWNR.G (Ions score 32)
385 - 392	452.2459	902.4773	902.4610	18	0 K.ASPVQWR.V (Ions score 52)
393 - 406	785.4494	1568.8843	1568.8484	23	0 R.VELTIEDISQVLR.K (Ions score 48)
411 - 422	507.9303	1020.7691	1020.7471	15	1 R.LRFIDIDEDR.S (Ions score 51)
432 - 443	801.3818	1600.7491	1600.7409	5	0 K.EYIQSQEYNDK.L (Ions score 36)
520 - 531	663.3760	1324.7395	1324.7211	12	0 R.QALALQAGQQR.Q (Ions score 56)
543 - 555	708.8656	1415.7166	1415.7078	6	0 K.VAANNSSQIHTK.H (Ions score 51)
543 - 555	716.8680	1431.7215	1431.7028	13	0 K.VAANNSSQIHTK.H Oxidation (M) (Ions score 106)
600 - 613	513.2744	1036.8014	1036.7896	8	0 K.LTIFEQAHSTPR.K (Ions score 28)
628 - 640	724.3667	1446.7189	1446.7103	6	0 K.STPVPVSGEYQQR.S (Ions score 36)
641 - 652	654.3262	1266.6378	1266.6204	14	0 R.SASPVVSGEYK.N (Ions score 41)
641 - 652	623.2242	1266.6507	1266.6204	24	0 R.SASPVVSGEYK.N (Ions score 58)
641 - 652	674.3103	1346.6061	1346.5867	14	0 R.SASPVVSGEYK.N Phospho (ST) (Ions score 81) ←
653 - 660	439.2546	876.4946	876.4818	15	0 K.NTIQGIK.R (Ions score 21)
704 - 713	561.7996	1121.5847	1121.5717	12	0 R.TISDYIPKAR.R (Ions score 47)
714 - 726	508.6109	1022.8118	1022.7780	22	1 R.NVFPVMSDYVQ.Q (Ions score 10)
732 - 740	491.7939	981.5732	981.5607	13	0 K.NTIAPPIR.S (Ions score 24)
746 - 763	989.5488	1977.0826	1977.0531	15	0 K.QNSLSPALQNAELVQK.Q (Ions score 80)
856 - 882	1032.1285	3093.3637	3093.3548	3	1 K.ERVNPSGNSFTVFEDSTTSGMDTNR.L (Ions score 62)
883 - 894	699.3809	1396.7073	1396.7086	-1	0 R.LRFVDSLEIK.Q (Ions score 62)
895 - 902	459.2582	916.5018	916.4865	17	0 K.QILEKSK.A (Ions score 12)
903 - 915	680.3405	1358.6664	1358.6540	9	0 K.AFPQSHPSIDYFK.S (Ions score 51)
903 - 915	688.3393	1374.6640	1374.6489	11	0 K.AFPQSHPSIDYFK.S Oxidation (M) (Ions score 27)
938 - 946	526.8238	1051.6330	1051.6138	18	0 R.NHISVILK.H (Ions score 52)
1023 - 1040	677.7172	2030.1289	2030.0769	26	1 K.RGQNNIPILPLATNQR.N (Ions score 66)
1023 - 1040	508.5410	2030.1347	2030.0769	28	1 K.RGQNNIPILPLATNQR.N (Ions score 14)
1081 - 1100	1066.0333	2130.0521	2130.0301	10	0 R.AQNHVNVGQEQVTFWTSQIK.E (Ions score 187)
1107 - 1113	443.2667	884.5188	884.5120	8	0 K.FEIHIVK.V (Ions score 27)
1116 - 1125	520.8239	1039.6332	1039.6179	15	0 R.IVGLAIVFPEK.K (Ions score 38)
1127 - 1135	534.2873	1066.5600	1066.5447	14	0 K.VSNTWLVK.N (Ions score 61)
1136 - 1143	468.7756	935.5367	935.5327	4	0 K.ELASYILK.E (Ions score 18)



De novo sequencing

De novo = peptide sequencing performed without prior knowledge of the amino acid sequence

- if enough material is available classical Edman degradation is still a very good method
- partial peptide sequencing is possible based on MS/MS data



Publishing protein ID data

- most biological journals do not have any guidelines yet

Mat+Met:

- MS instrument, identification type
- data acquisition+processing software
- database search engine

Results: Table from ID results

Protein name	Fold difference	Protein number	Access no.		Identification type	No. of peptide matched	Sequence coverage	Mowse score	Theoretical		Biological process
			SwissProt	NCBItr					MW	pl	
Upregulated proteins											
14-3-3 protein sigma (epithelial cellmarker 1 / stratifin)	x	73*	P31947	187302	LC-MS/MS	10	32 %	216	27873	4.72	Signaling
cytokeratin 13 (frag.)			P13646	34033	LC-MS/MS	11	22 %	142	48666	4.87	Structural
14-3-3 protein sigma (epithelial cellmarker 1 / stratifin)	x	74*	P31947	187302	LC-MS/MS	12	39 %	444	27873	4.72	Signaling
cytokeratin 13 (frag.)			P13646	30377	LC-MS/MS	11	22 %	372	46179	4.83	Structural

Tutorial for Proteomics Data Submission

Katalin F. Medzihradzky
Robert J. Chalkley
UCSF

Mol Cell Proteomics
www.mcponline.org



Why Have Guidelines?

- Large-scale proteomics studies create huge amounts of data.
- It is impossible/impractical to present all the results.
- The function of the guidelines is to:
 - Provide enough information to be able to explain the experiment.
 - Provide an assessment of the reliability of the results.
 - Provide the data that supports the results, particularly those that have the greatest potential for mis-interpretation, so the readers can manually assess the results that are important to them.

The Problem

- Proteomics experiments are carried out by many different methods, using a variety of instrument types and employing different analysis tools. Hence, many experimental and analysis parameters need to be reported and the parameters required will differ depending on how the experiment was performed and analyzed.



In the experimental section:

Peak Picking Software

- In the **Experimental** Section
- **Name of peaklist-generating software and release version (number or date)**
 - The raw data acquired by the mass spectrometer is converted into a centroided peaklist file for database searching. The program that performs this can do many things, such as:
 - Remove peaks that are below a certain intensity or signal to noise ratio;
 - Require peaks to have a certain resolution in order to make the list;
 - Set a threshold for the maximum number of peaks within a mass range;
 - Assign a charge state to ions;
 - Merge together MS/MS spectra that have the same precursor mass.

Peak Picking (cont'd)

In the **Experimental** Section

- **Parameters used – default vs altered**

- Using different peak picking software / parameters will change the subsequent database searching results, so it is necessary to report the software used for peaklist generation and parameters used.

Examples of software

- Extract_msn in Bioworks 3.0 (Thermo)
- Mascot.dll v1.6b19 (Applied Biosystems)
- DataAnalysis 3.2 (Bruker)
- Mascot Distiller v2.1

- As many people use default parameters for whatever software they use, by specifying the version of the software it can be sufficient to state that the default parameters were used.

Search Engine

In the **Experimental** Section

- **Name of the search engine and release version (number or date)**

- Search engines change over time:
 - An improved scoring system may be implemented.
 - New ways to filter the results may be included.

Hence, the version number of the search engine is important.

Database Search Parameters

In the **Experimental** Section

- Enzyme specificity considered
- # of missed cleavages permitted
- Fixed modification(s) (including residue specificity)
- Variable modification(s) (including 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.

Single-Peptide-Based Protein IDs and PTMs

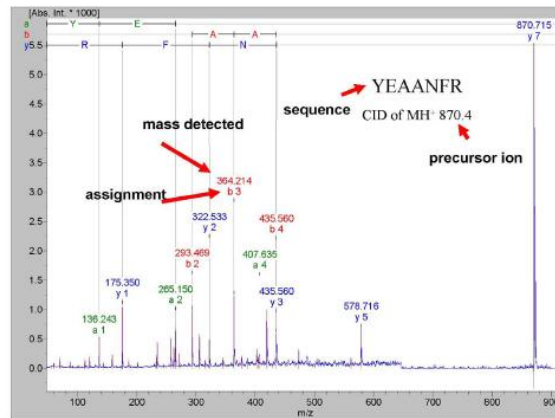
In the **Results** Section

- For each protein or site, a MS/MS spectrum appropriately labeled should be included, with masses detected as well as fragment assignments
- It is recognized that properly labeled spectra are not always readily produced.
- Other acceptable options (in order of preference):
 - Two copies of the same spectrum/page – one labeled with the masses, the other with fragment assignments;
 - Spectrum labeled with masses, accompanied with a Table of the fragment assignments;
 - Fragment assignments provided by the search engine in the spectrum and corresponding masses highlighted in a Table of theoretical fragments (e.g. Mascot results output).

These files can be quite large. If there are large numbers of spectra it is often easier to split them between multiple files. We suggest to not include data on more than 50 proteins in a single file.

Example Presentation of Spectrum of a Single-Peptide-Based Protein ID

In the Results Section



The Future

- These guidelines will be continuously adjusted to the ever changing needs of the proteomics field.
- As software improves to become easier to output results into compliant formats, it will become easier to properly present proteomics data in any forum.
- As common formats e.g. mzML, AnalysisXML, become universally available, tools will be produced that will be able to automatically extract the relevant information from raw data and search results.
- Once these formatting problems are solved, journal submission guidelines for proteomics data are expected to become similar for all journals.
- Making raw data publicly available is going to become more common.

The screenshot displays the PRIDE Archive website in a browser window. At the top, there is a cookie notice and a search bar. The main header includes the EMBL-EBI logo and the PRIDE Archive title. Below the header, the page is organized into several sections:

- PRIDE Archive - proteomics data repository:** A central heading with a brief description of the database as a centralized, standards-compliant, public data repository for proteomics data.
- Submit data:** A section for users to submit their MS/MS proteomics datasets.
- Access data:** A section for browsing public datasets, including links to BioMart and ProteomeXchange.
- Tools:** A section listing software tools like PRIDE Inspector and PRIDE Converter.
- News:** A section for recent news items, such as "PKD000592: Mouse Core Acrossomal Matrix LC-MS/MS link in Mol Cell Biol." and "PKD000807: Human oral cancer brush biopsy ITRAQ link in PLoS One".

At the bottom of the page, there is a navigation menu with categories like Services, Research, Training, Industry, and About us. The footer contains copyright information and links to privacy and terms of use.

<http://www.ebi.ac.uk/pride/archive/>