MS/MS Database Searching

Jimmy Eng

Day 1

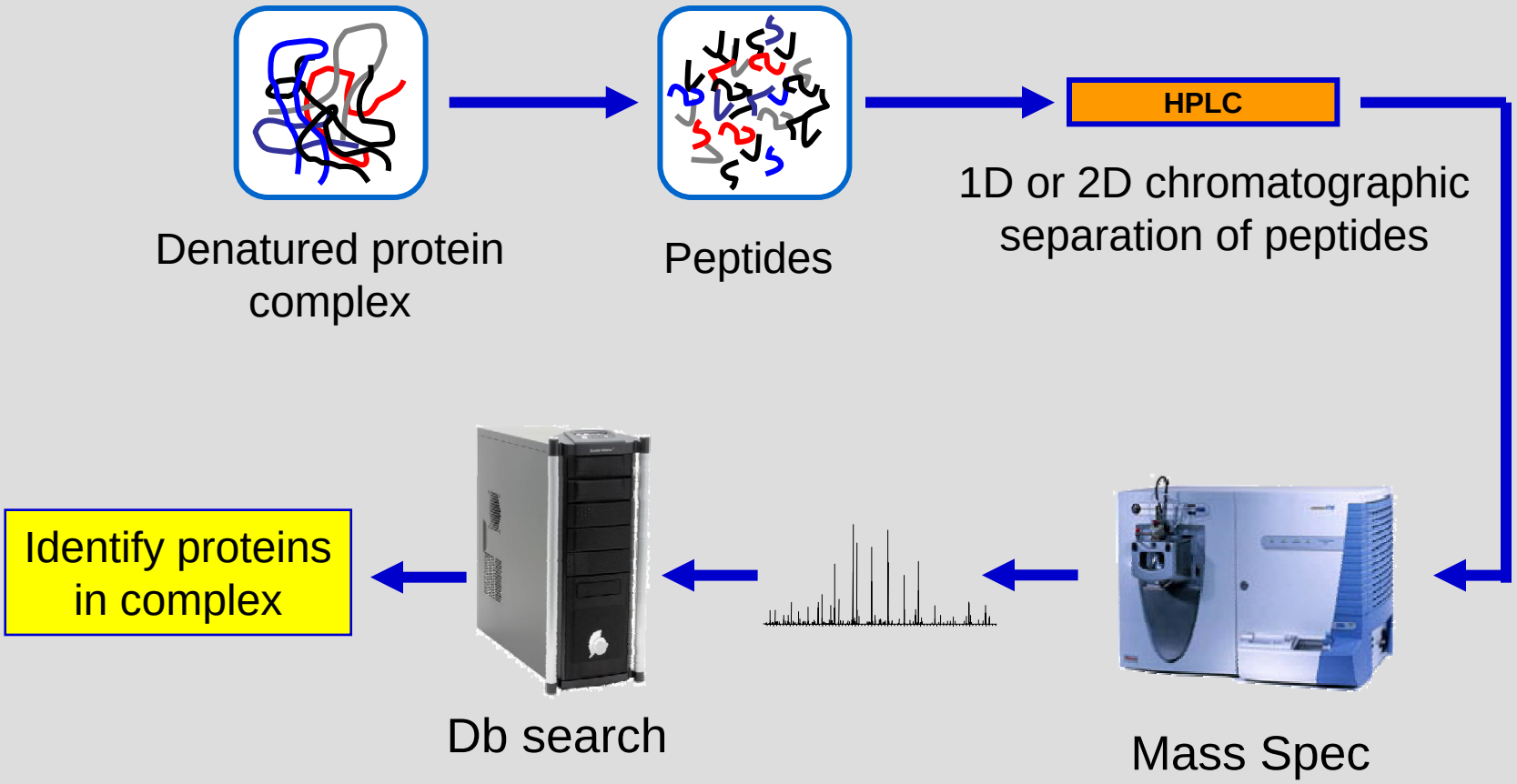
October 16, 2006

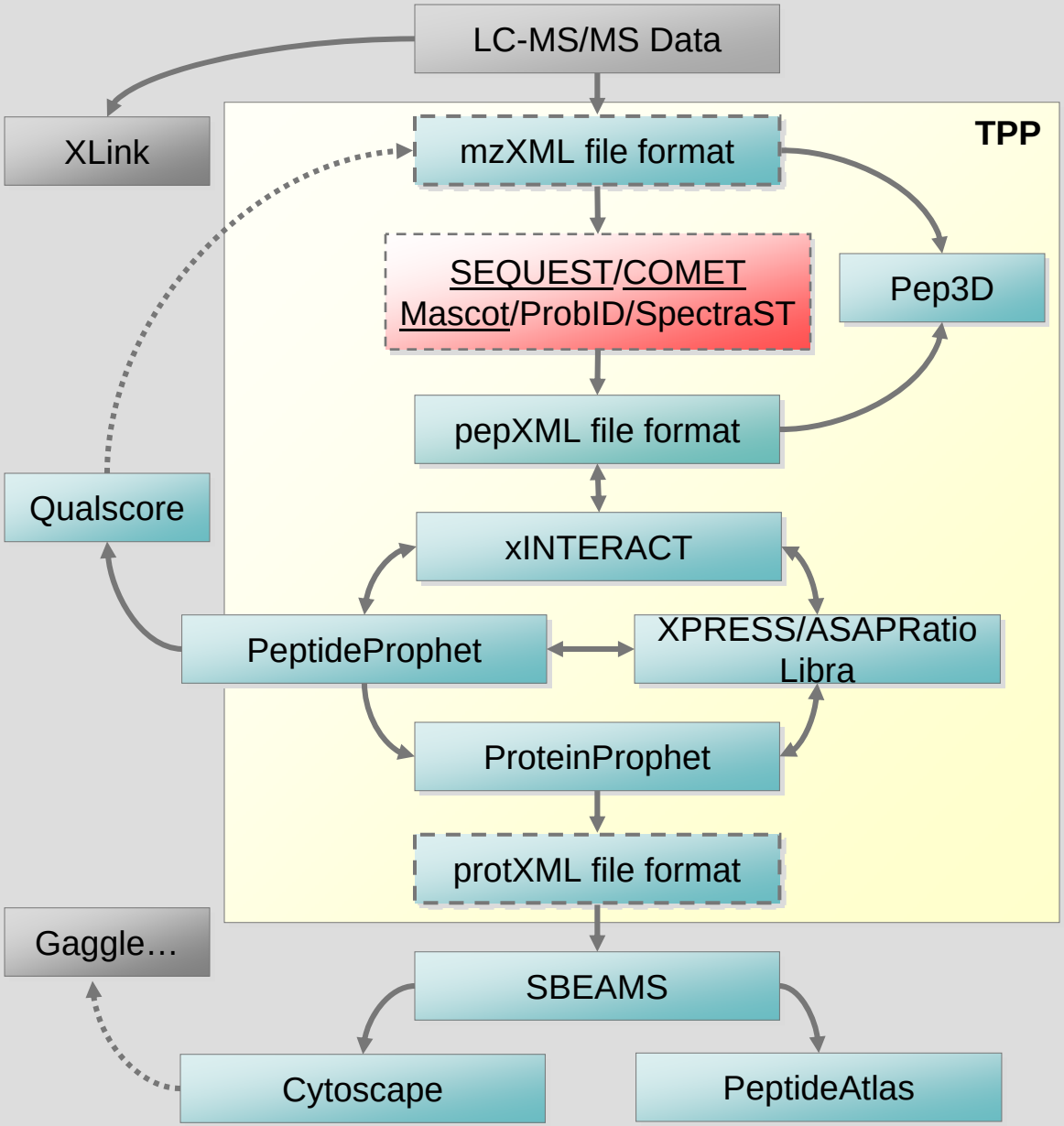
INSTITUTE FOR Systems Biology

Day 1 Lecture Topics

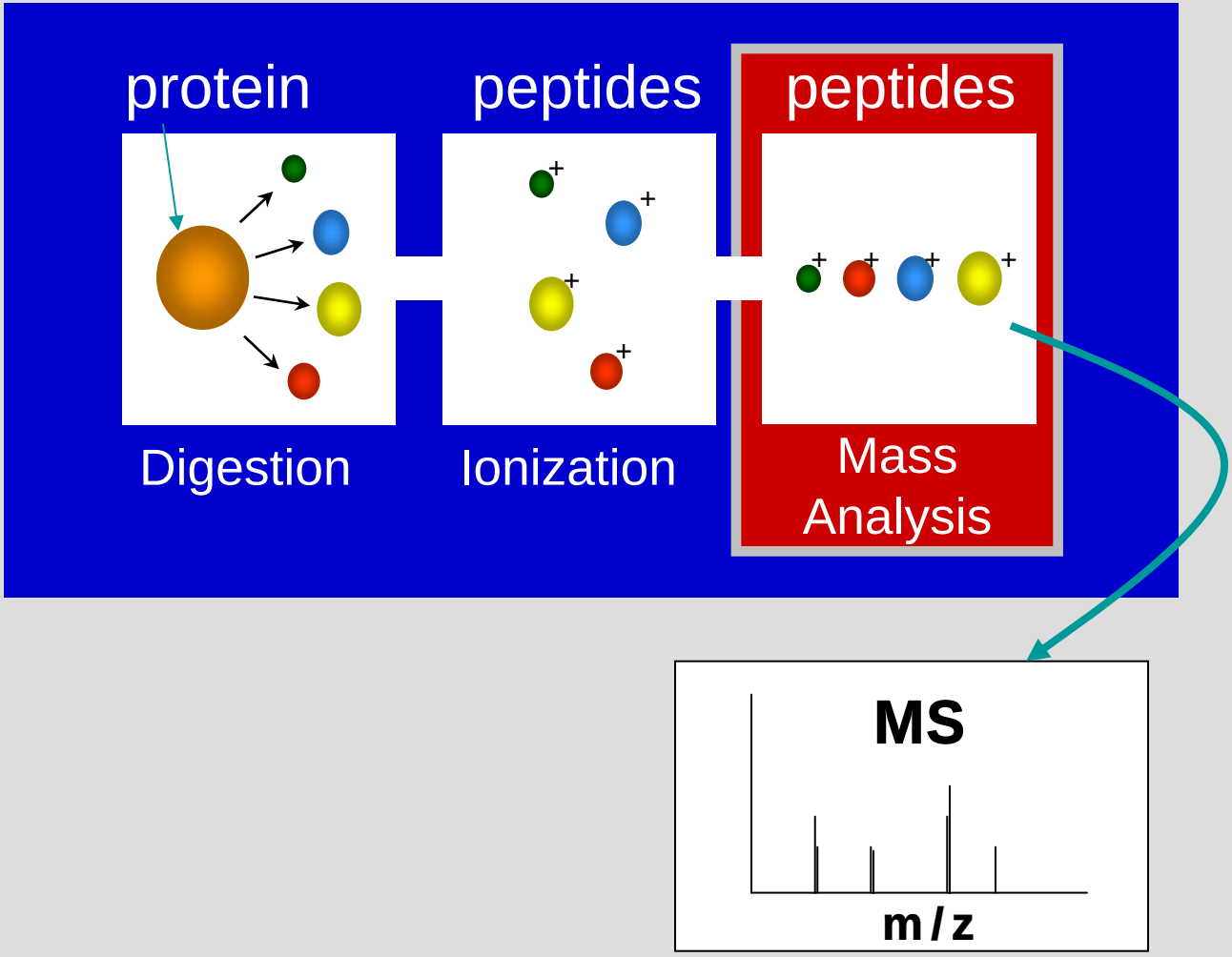
- Basic background & motivation
- Peptide fragmentation, nomenclature
- Peptide vs. tandem mass spectra
- Sequence database searching
  - Databases
  - Enzymes
  - Modifications
- Interpretation of search results; manual validation
- Introduction to software tools

Protein Identification Strategy

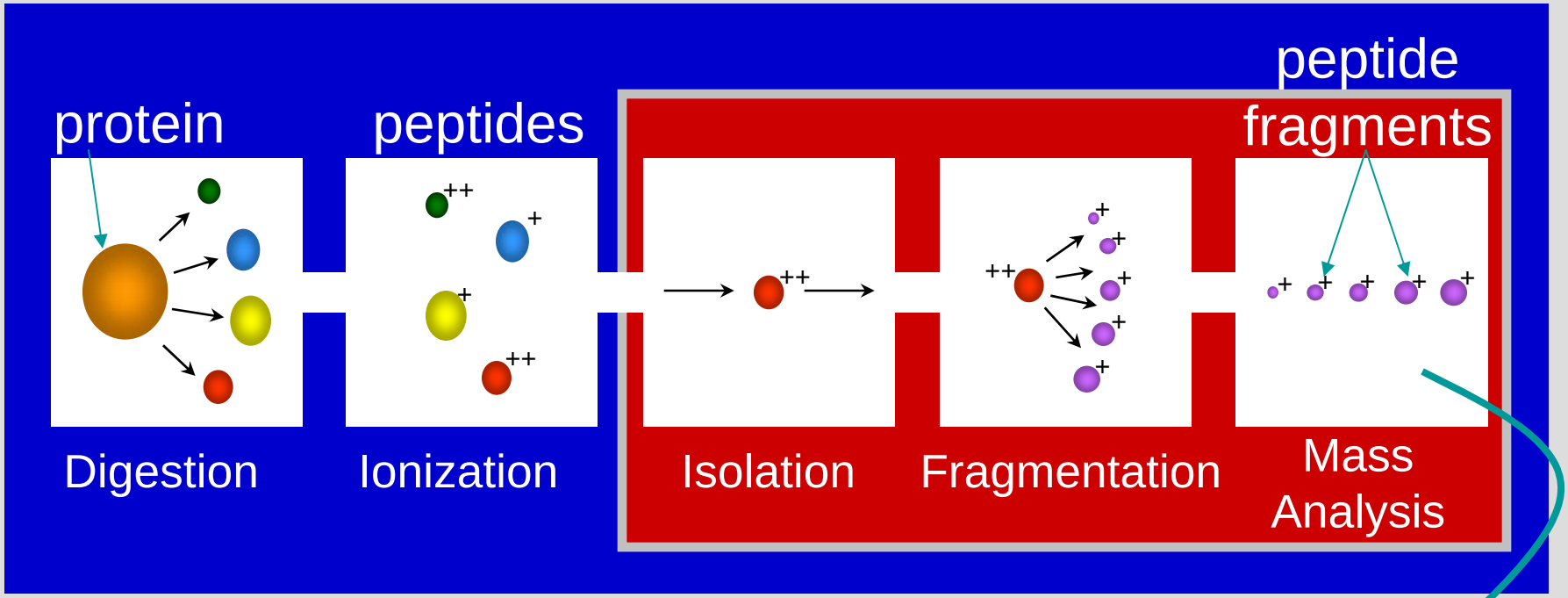


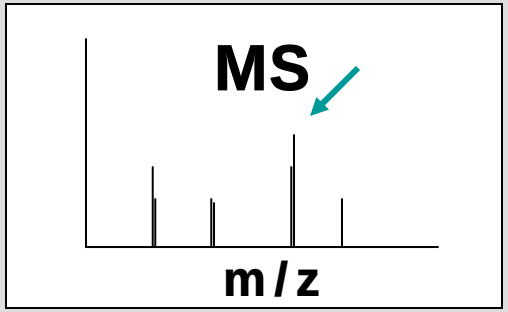

LC-MS/MS Data
XLink
TPP
mzXML file format
SEQUEST/COMET
Mascot/ProbID/SpectraST
Pep3D
pepXML file format
Qualscore
xINTERACT
PeptideProphet
XPRESS/ASAPRatio
Libra
ProteinProphet
protXML file format
Gaggle…
SBEAMS
Cytoscape
PeptideAtlas

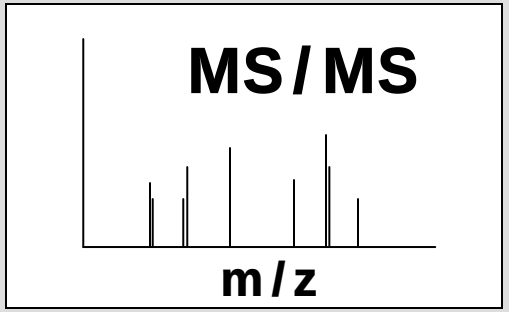
Single Stage MS



Tandem MS

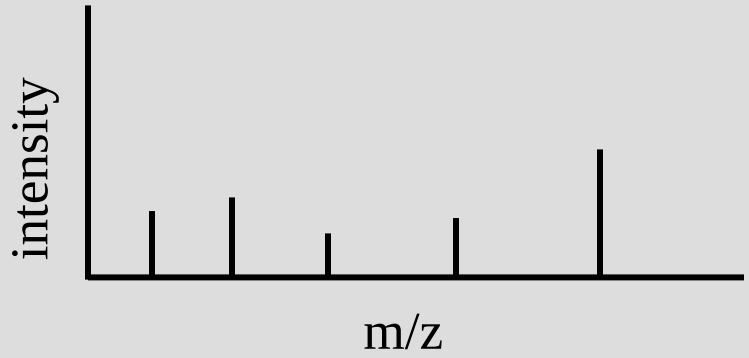

protein
peptides
peptide fragments
Digestion
Ionization
Isolation
Fragmentation
Mass Analysis


MS
m/z

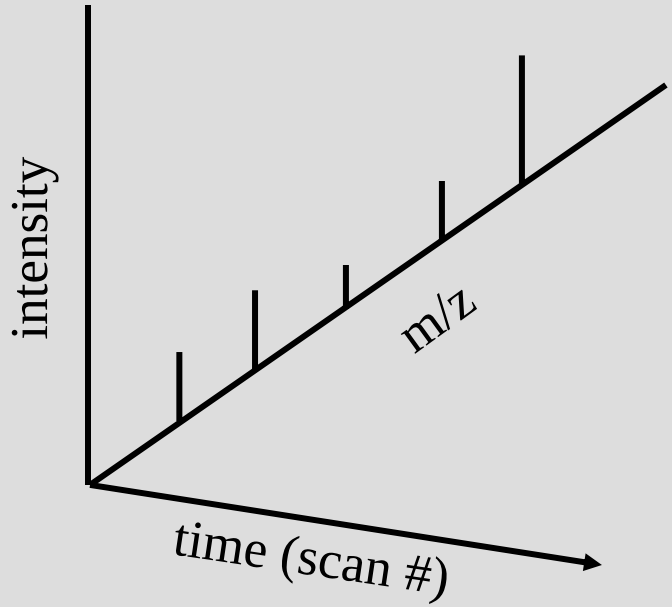

MS/MS
m/z

Mass vs. Intensity vs. Time

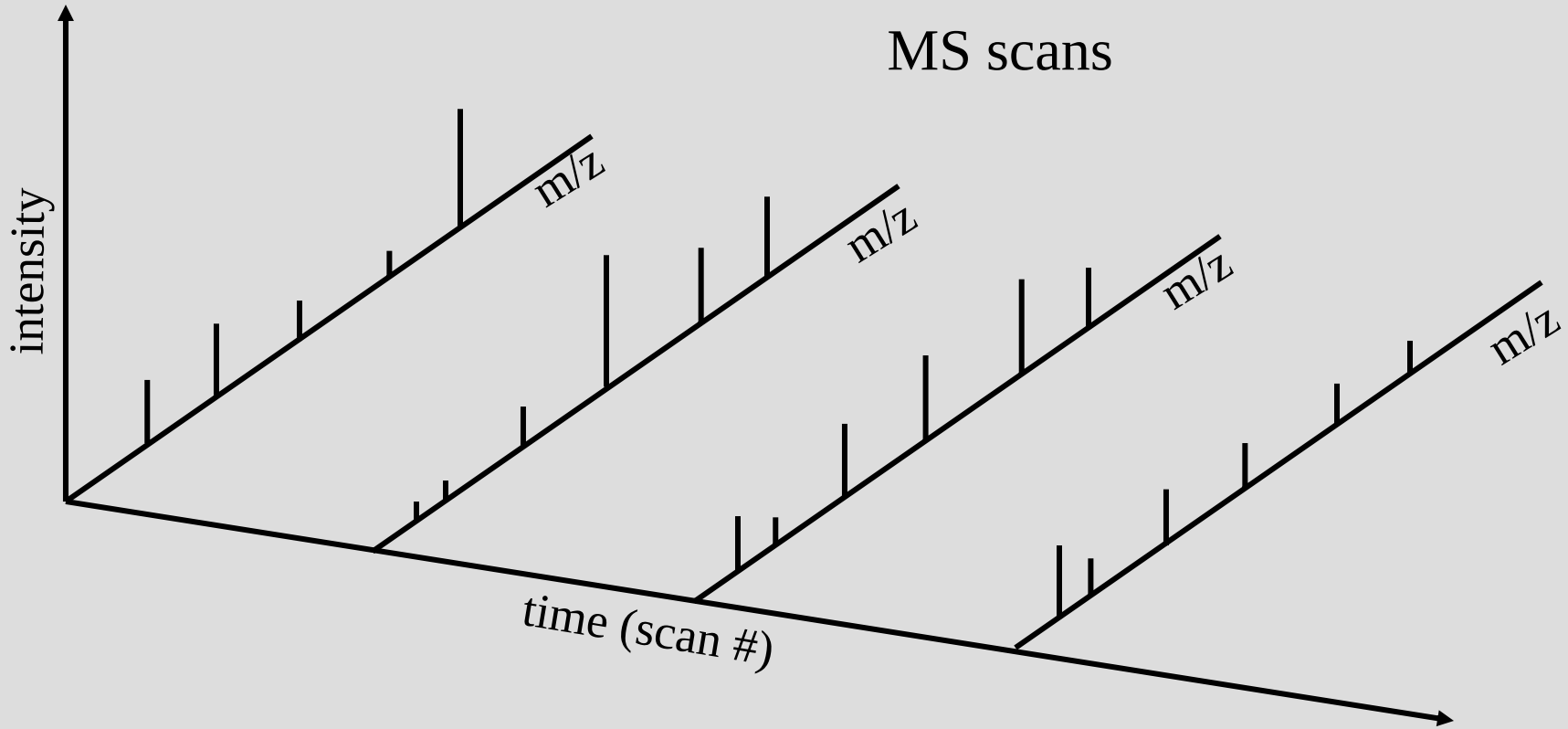
2D view: m/z, intensity



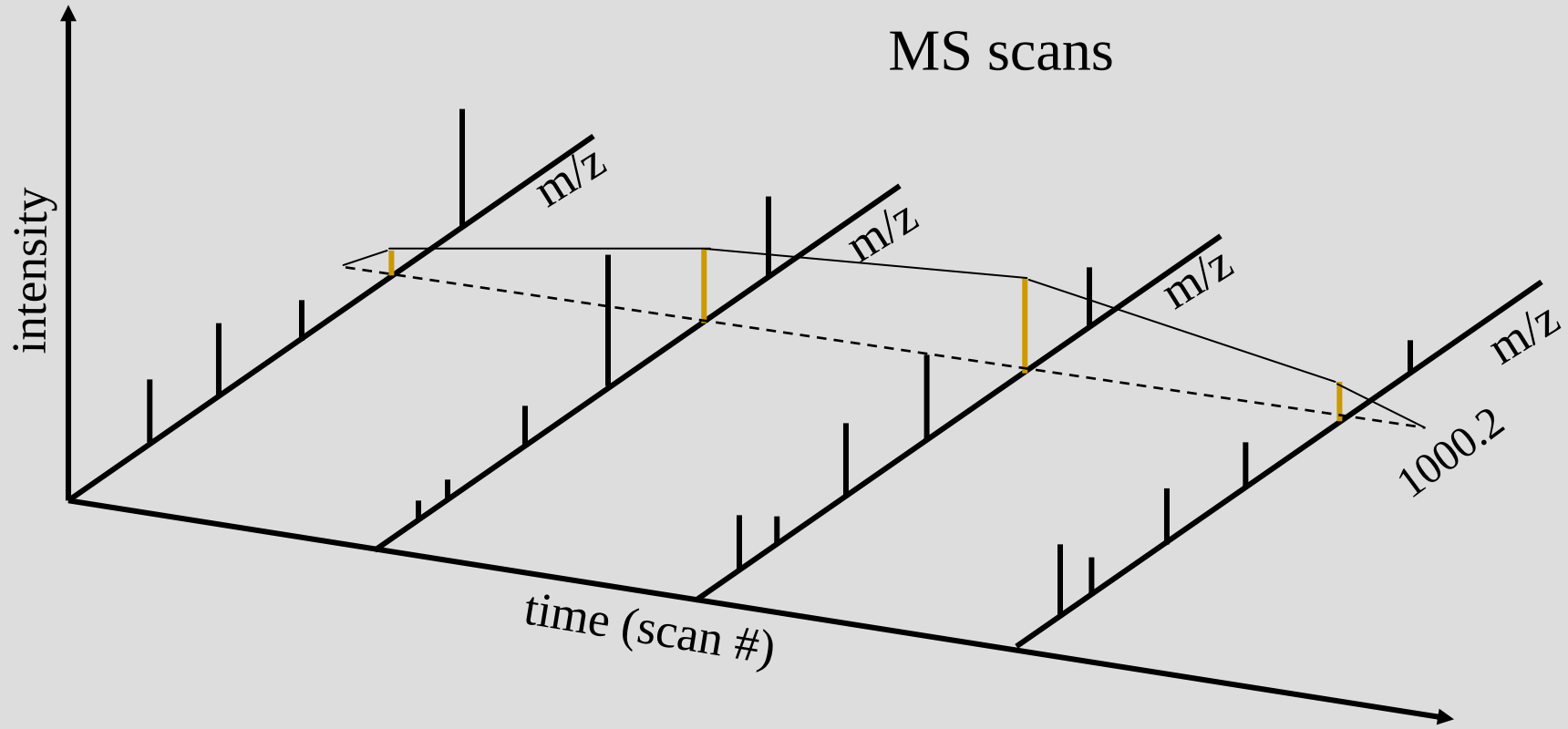
3D view: m/z, intensity, time



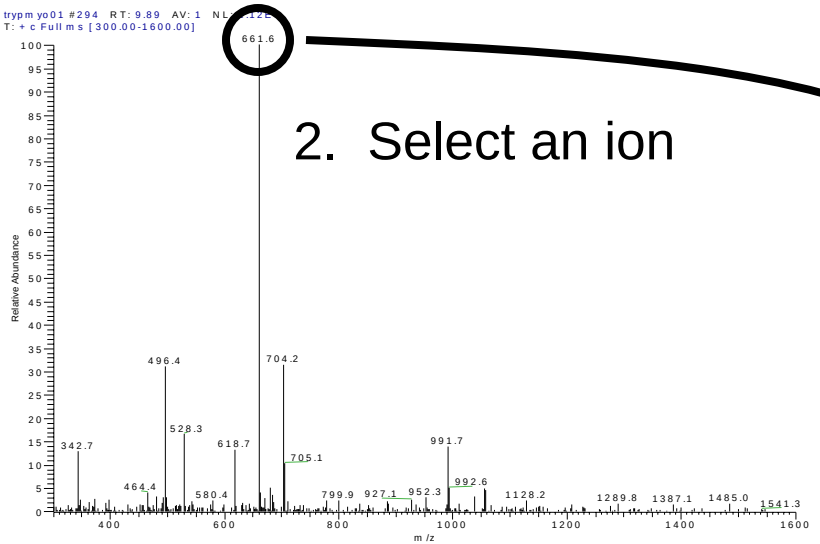
Mass vs. Intensity vs. Time



Mass vs. Intensity vs. Time



MS/MS Data Acquisition

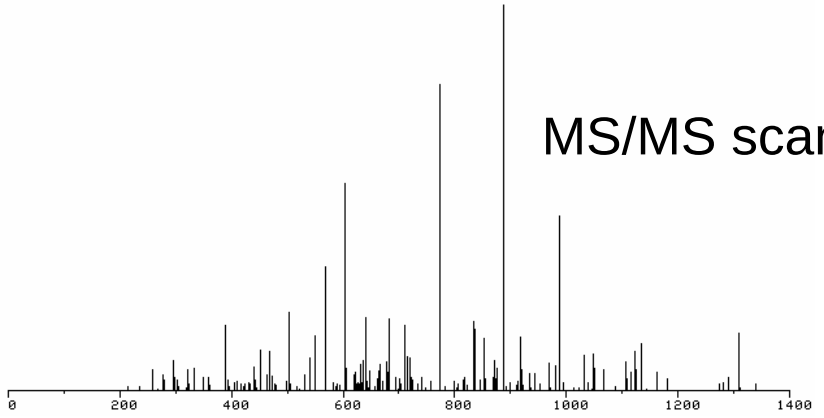


1. Acquire full (MS) scan

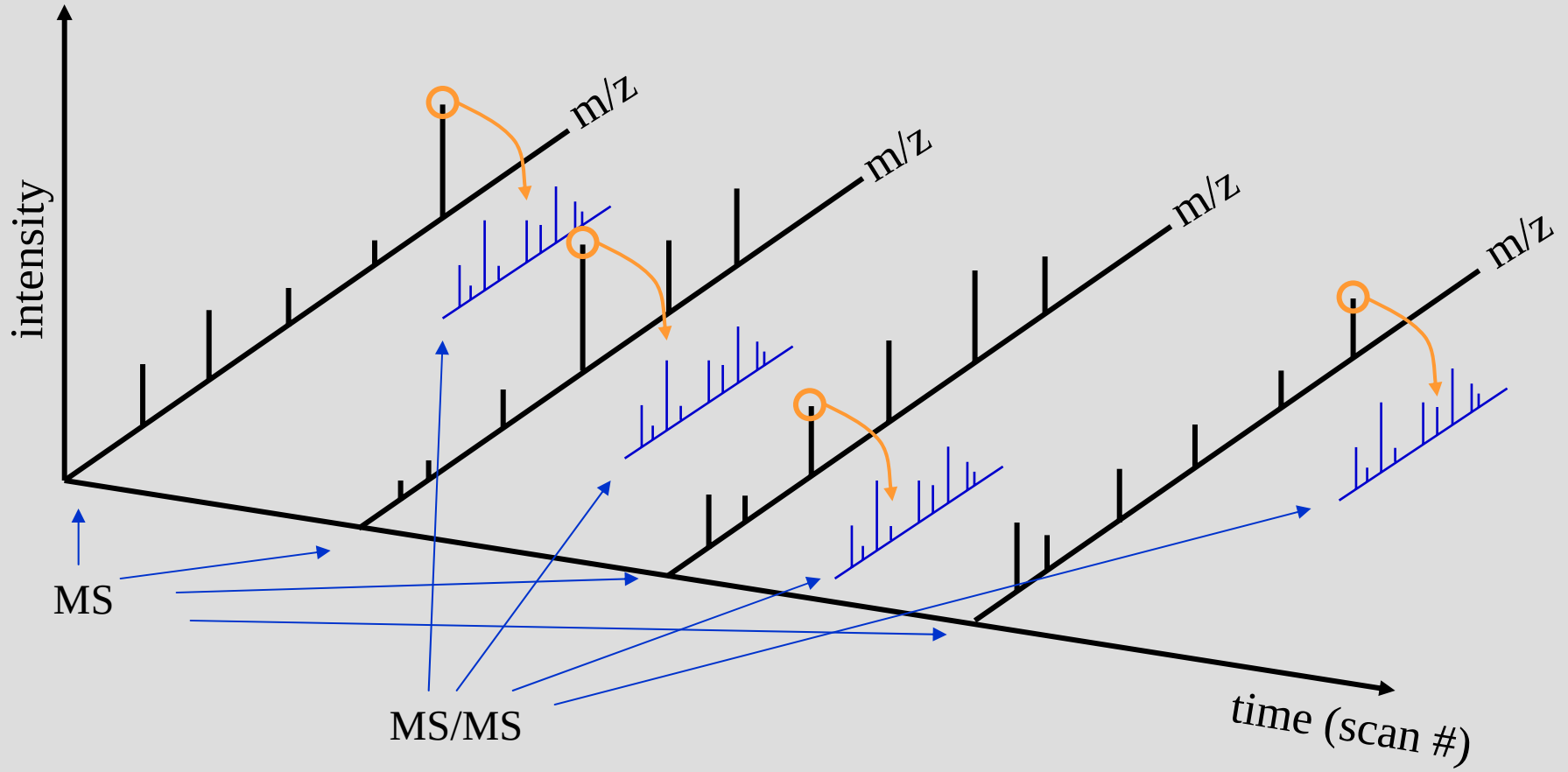
2. Select an ion

3. Isolate ion

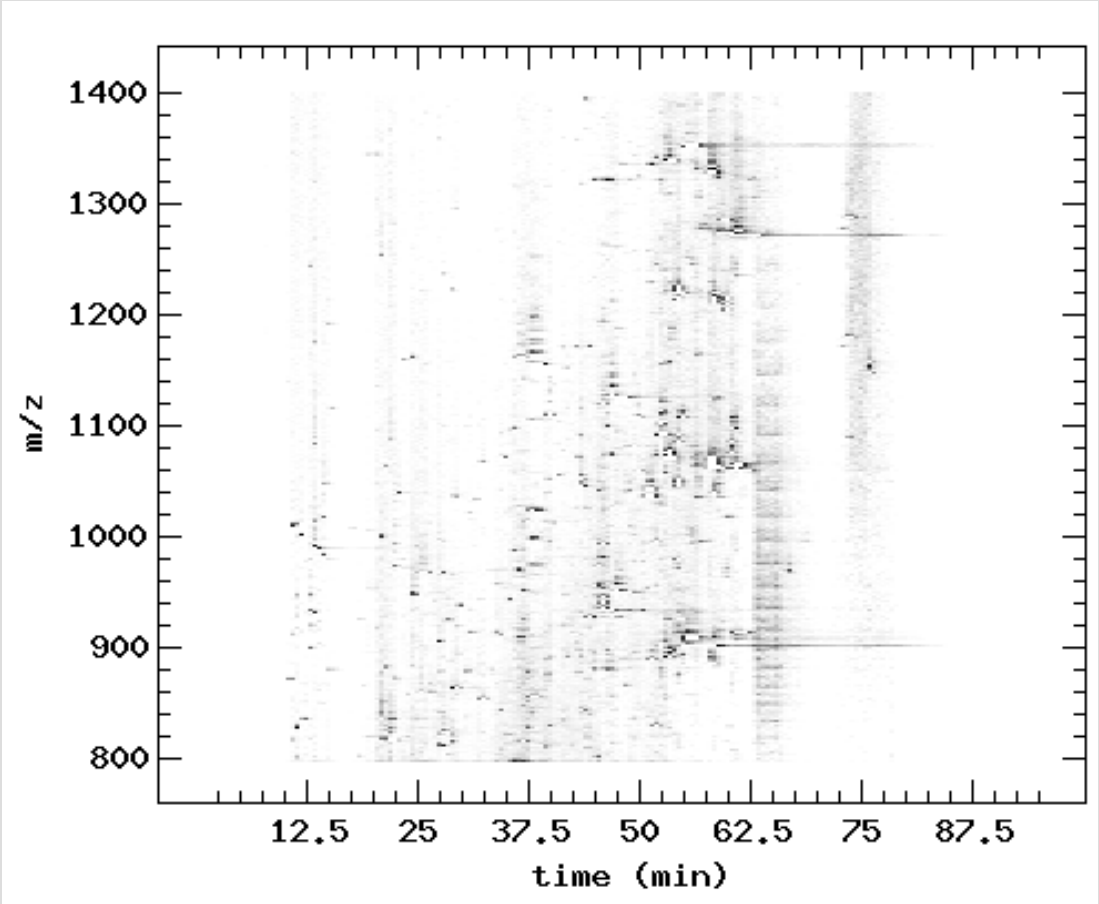
4. Fragment ion



MS vs. MS/MS



2D view of an LC-MS experiment



You'll learn all about Pep3D soon!

Amino Acids

Amino acid	3LC	SLC	Average	Monoisotopic
Glycine	Gly	G	57.0519	57.02146
Alanine	Ala	A	71.0788	71.03711
Serine	Ser	S	87.0782	87.02303
Proline	Pro	P	97.1167	97.05276
Valine	Val	V	99.1326	99.06841
Threonine	Thr	T	101.1051	101.04768
Cysteine	Cys	C	103.1388	103.00919
Leucine	Leu	L	113.1594	113.08406
Isoleucine	Ile	I	113.1594	113.08406
Asparagine	Asn	N	114.1038	114.04293
Aspartic acid	Asp	D	115.0886	115.02694
Glutamine	Gln	Q	128.1307	128.05858
Lysine	Kys	K	128.1741	128.09496
Glutamic acid	Glu	E	129.1155	129.04259
Methionine	Met	M	131.1926	131.04049
Histidine	His	H	137.1411	137.05891
Phenyalanine	Phe	F	147.1766	147.06841
Arginine	Arg	R	156.1875	156.10111
Tyrosine	Tyr	Y	163.1760	163.06333
Tryptophan	Trp	W	186.2132	186.07931

Average vs. Monoisotopic Mass

For example:

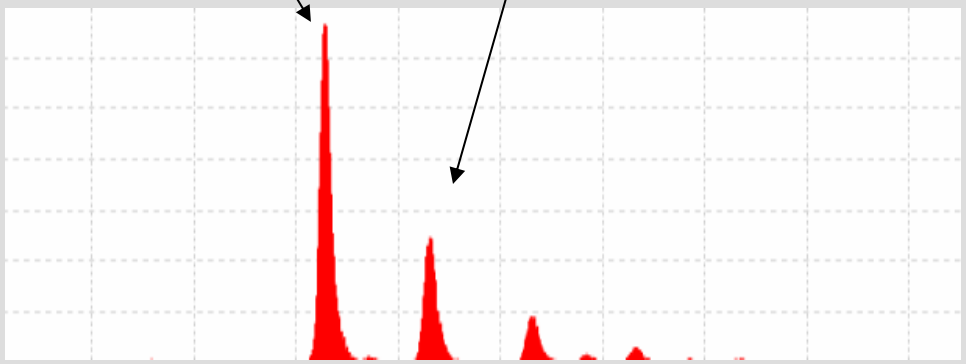
DIGSESTEDQAMEDIK

Mono MH+: 1767.7594 Da

Avg MH+: 1768.8438 Da

Monoisotopic mass

Average mass – centroid of isotopic envelope

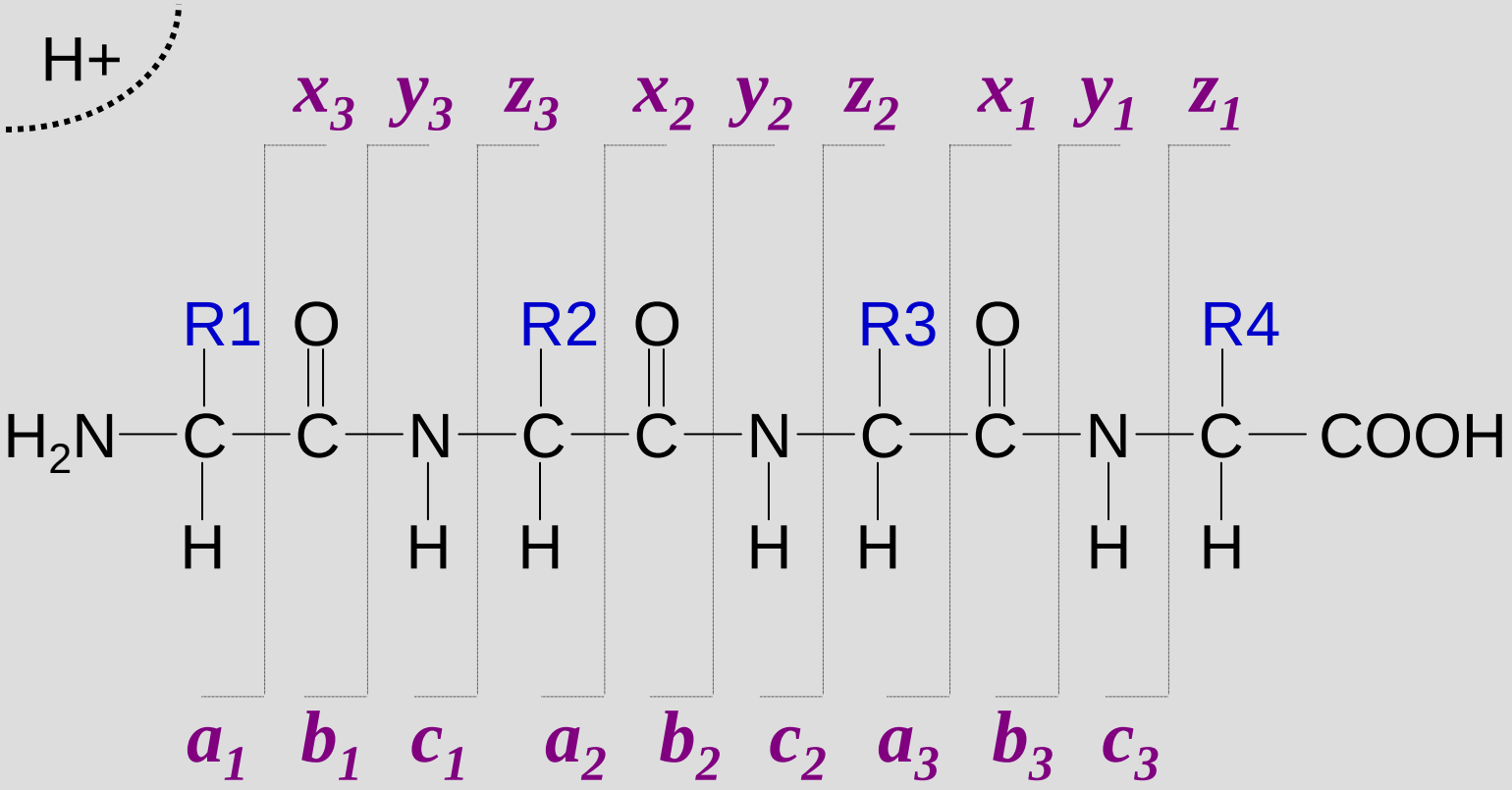


Δm

Difference in mass can be significant!

Charge state = $1 / \Delta m$

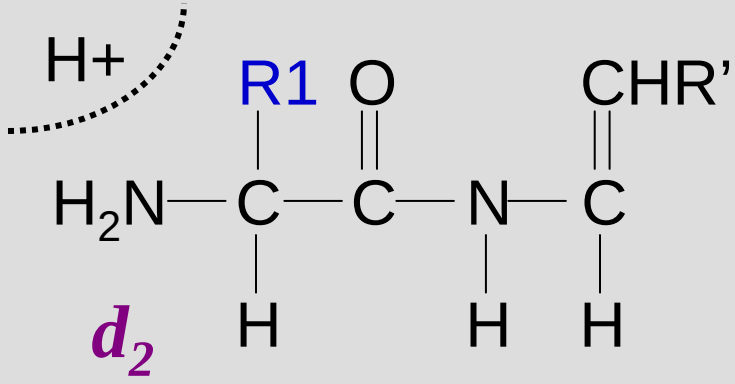
Fragment Ions



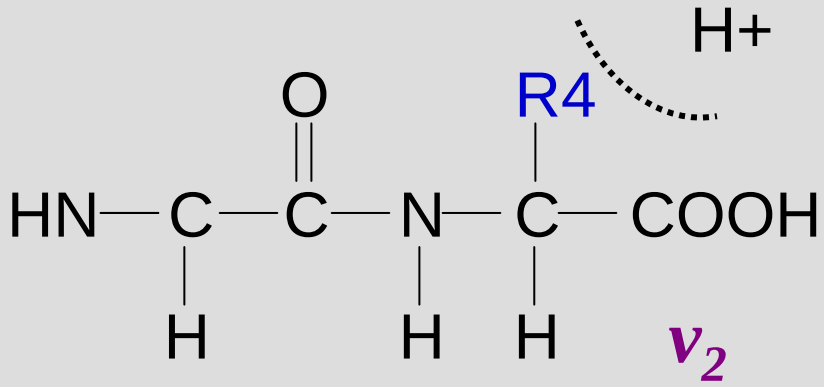
http://www.matrixscience.com/help/fragmentation_help.html

Fragment Ion Types

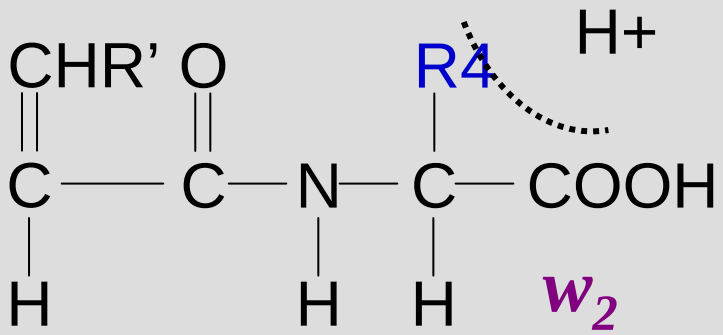
d-, *v*-, and *w*-ions are created by side chain cleavage. These ions are typically generated during high energy collision induced dissociation conditions. Of note, *d*- and *w*- ions allow the isobaric residues leucine and isoleucine to be differentiated.



d_2



v_2



w_2

http://www.matrixscience.com/help/fragmentation_help.html

Immonium Ions

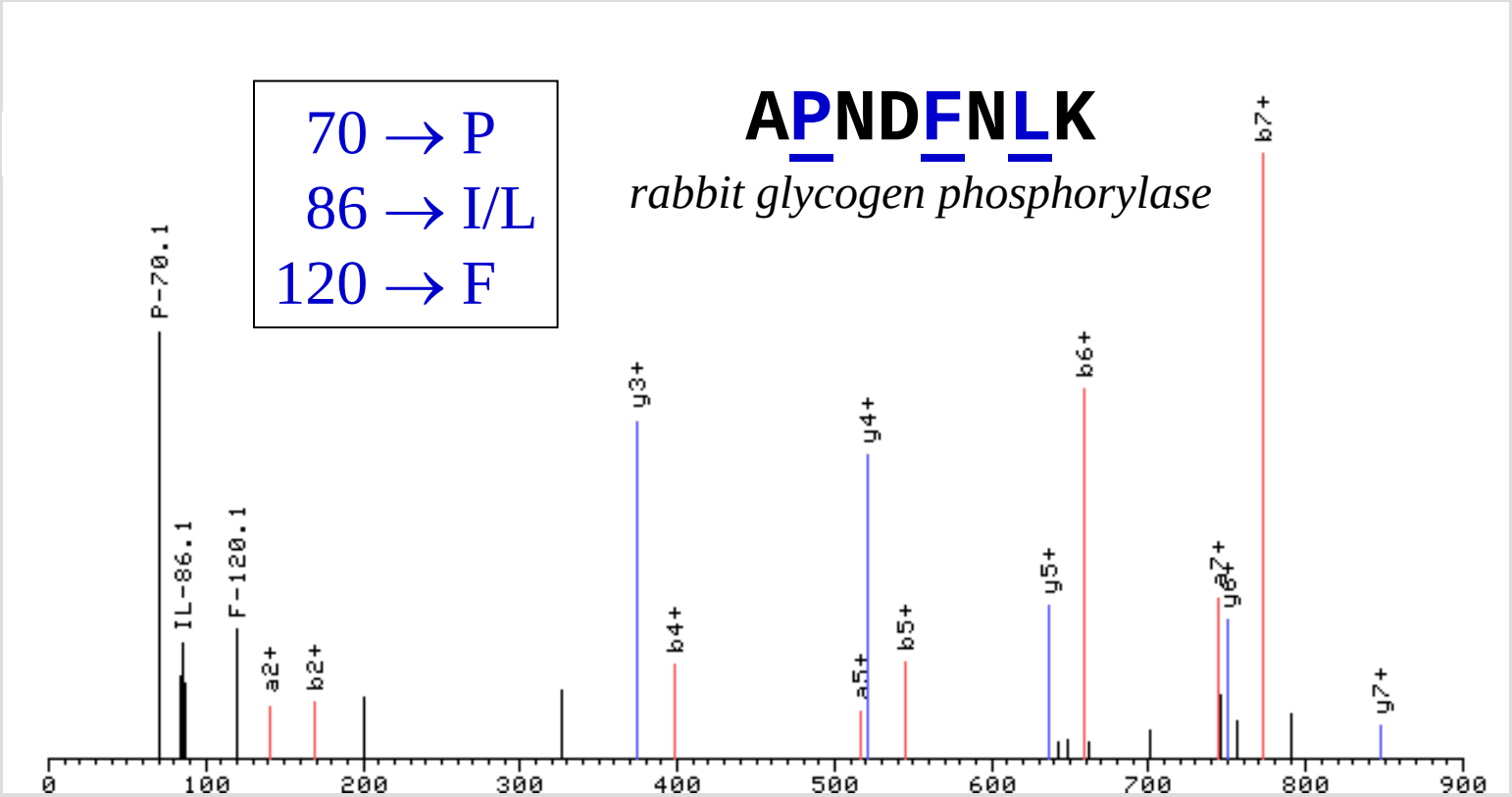
An internal fragment with just a single side chain formed by a combination of *a* type and *y* type cleavage is called an immonium ion. The presence of these ions can be a diagnostic to the presence of the corresponding amino acid in the peptide sequence.

Amino Acid	Residue Mass	Immonium ion mass	
Glycine	57.02147	30.03438	-
Alanine	71.03712	44.05003	-
Serine	87.03203	60.04494	+
Proline	97.05277	70.06568	++
Valine	99.06842	72.08133	++
Threonine	101.04768	74.06059	+
Cysteine	103.00919	76.0221	-
- carbamidomethylated	160.03065	133.0436	+
- carboxymethylated	161.01466	134.0276	+
- acrylamide adduct	174.0643	147.0772	+
Isoleucine	113.08407	86.09698	++
Leucine	113.08407	86.09698	++

Amino Acid	Residue Mass	Immonium ion mass	
Asparagine	114.04293	87.05584	+
Aspartic acid	115.02695	88.03986	+
Glutamine	128.05858	101.0715	+
Lysine	128.09497	101.1079 (84.08136)	
Glutamic acid	129.0426	102.0555	+
Methionine	131.04049	104.0534	+
- oxidized methionine	147.0354	120.0483	+
Histidine	137.05891	110.0718	++
Phenylalanine	147.06842	120.0813	++
Arginine	156.10112	129.114	-
Tyrosine	163.06333	136.0762	++
Tryptophan	186.07392	159.0922	+

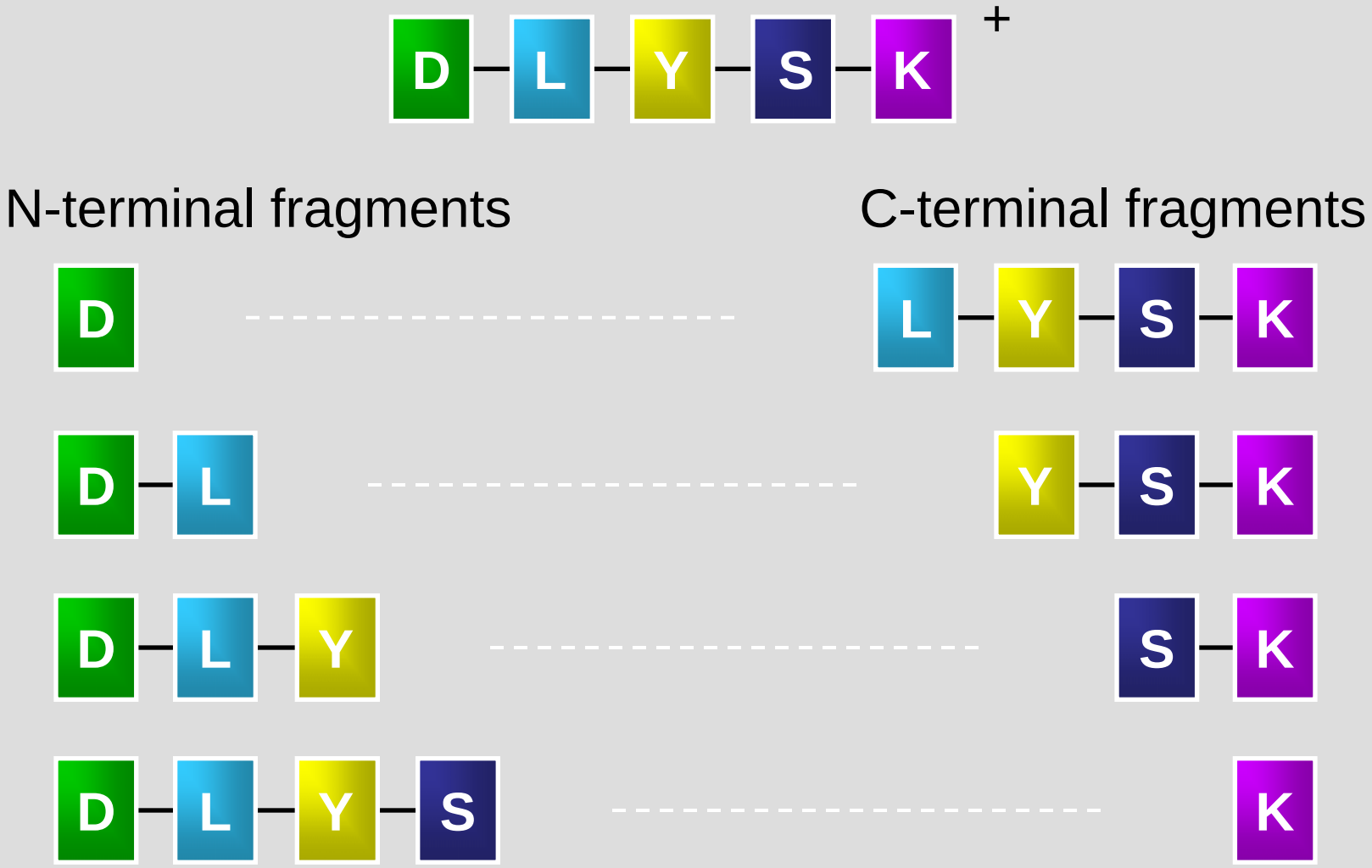
http://www.matrixscience.com/help/fragmentation_help.html
http://www.abrf.org/ResearchGroups/MassSpectrometry/EPosters/ms97quiz/residueMasses.html

Immonium Ions



MALDI-TOF-TOF tandem mass spectrum

Peptide Fragmentation



Fragmenting a Peptide

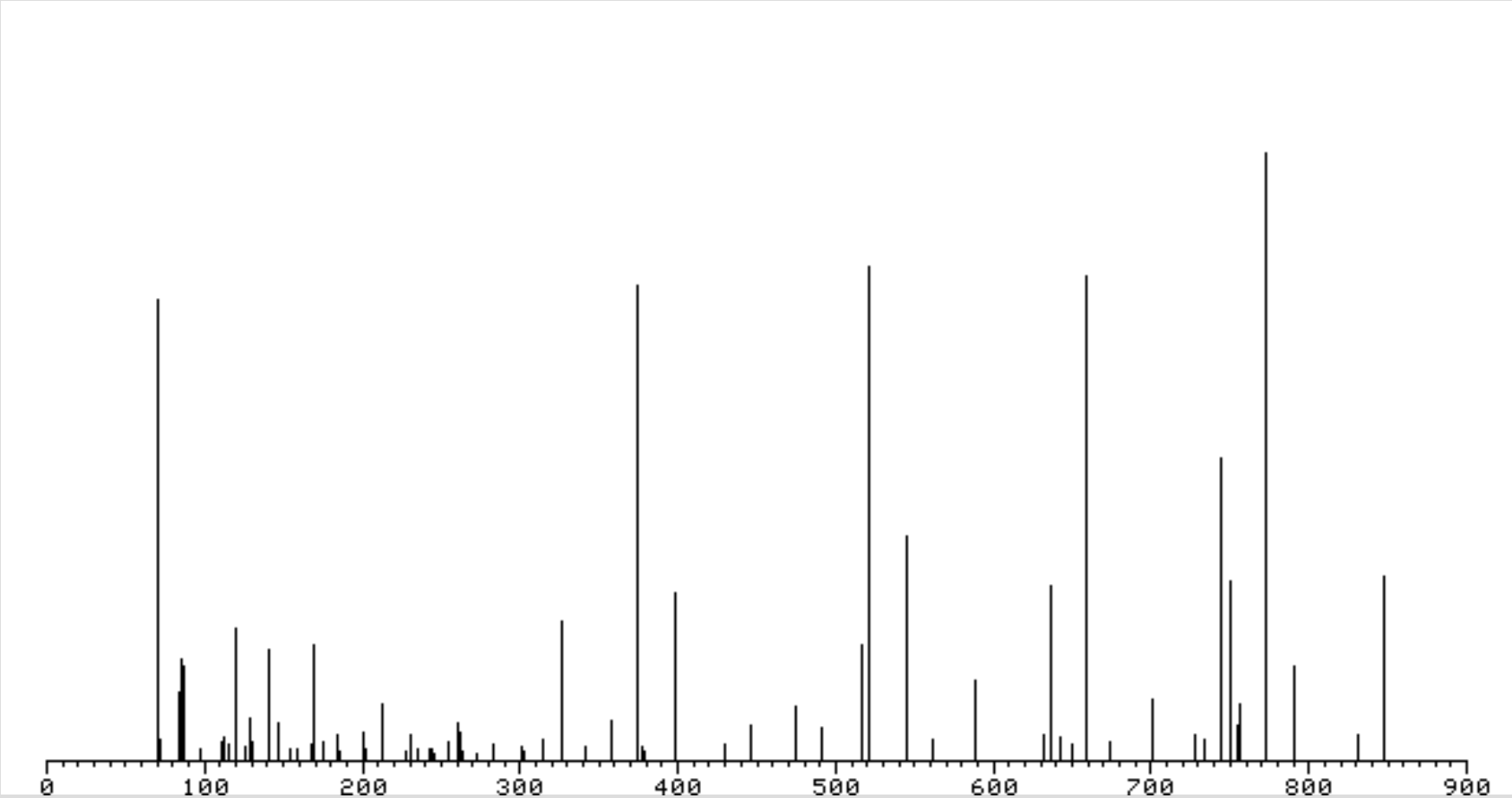
A-P-N-D-F-N-L-K

(MH+ 918.5)

B-ions		Y-ions
72.0	A ..... P-N-D-F-N-L-K	847.4
169.1	A-P ..... N-D-F-N-L-K	750.4
283.1	A-P-N ..... D-F-N-L-K	636.3
398.2	A-P-N-D ..... F-N-L-K	521.3
545.2	A-P-N-D-F ..... N-L-K	374.2
659.3	A-P-N-D-F-N ..... L-K	260.2
772.4	A-P-N-D-F-N-L ..... K	147.1

monoisotopic masses

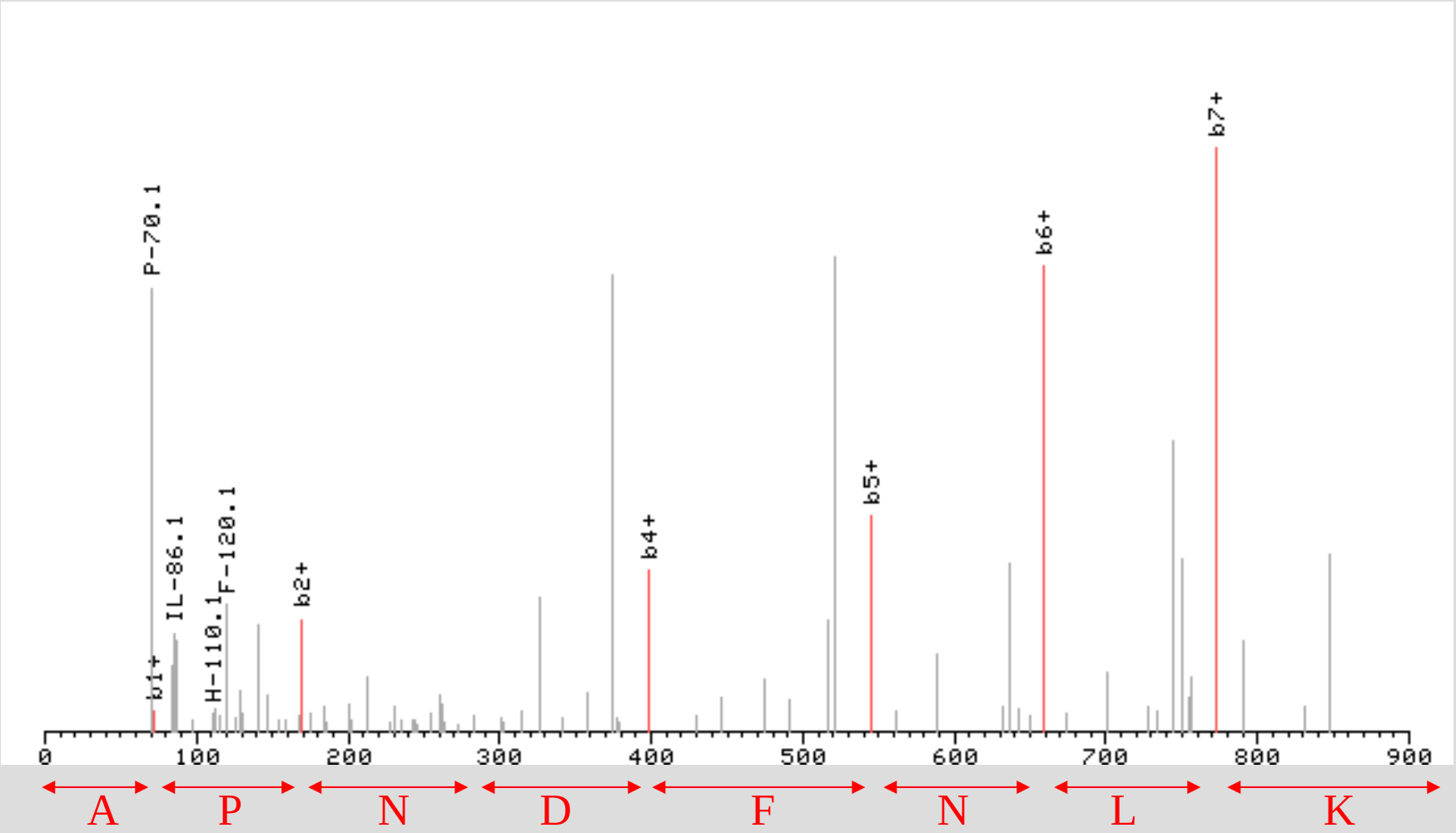
Sequence vs. Tandem Mass Spectrum



A-P-N-D-F-N-L-K

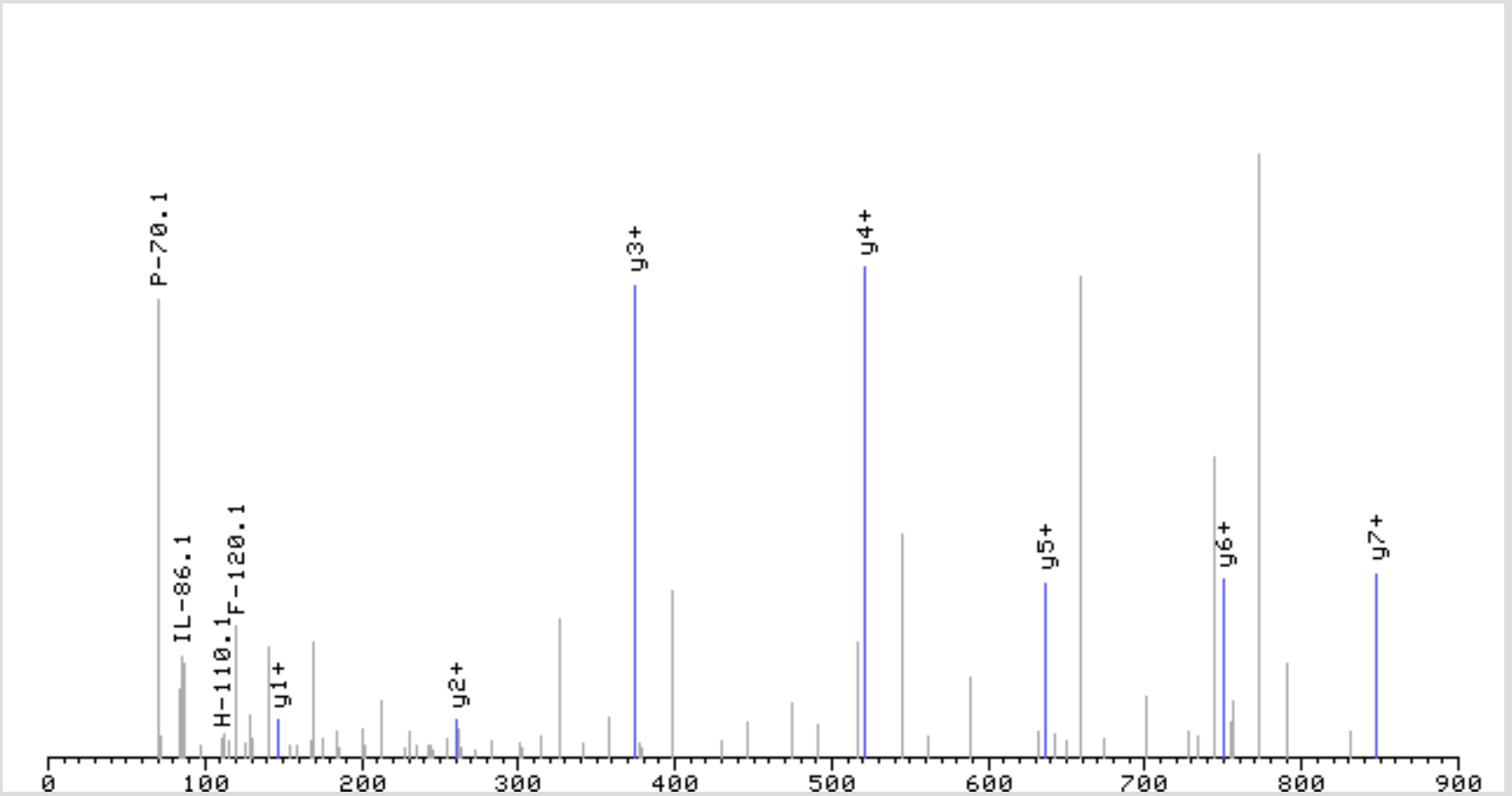
(MH+ 918.5)

Sequence vs. Tandem Mass Spectrum



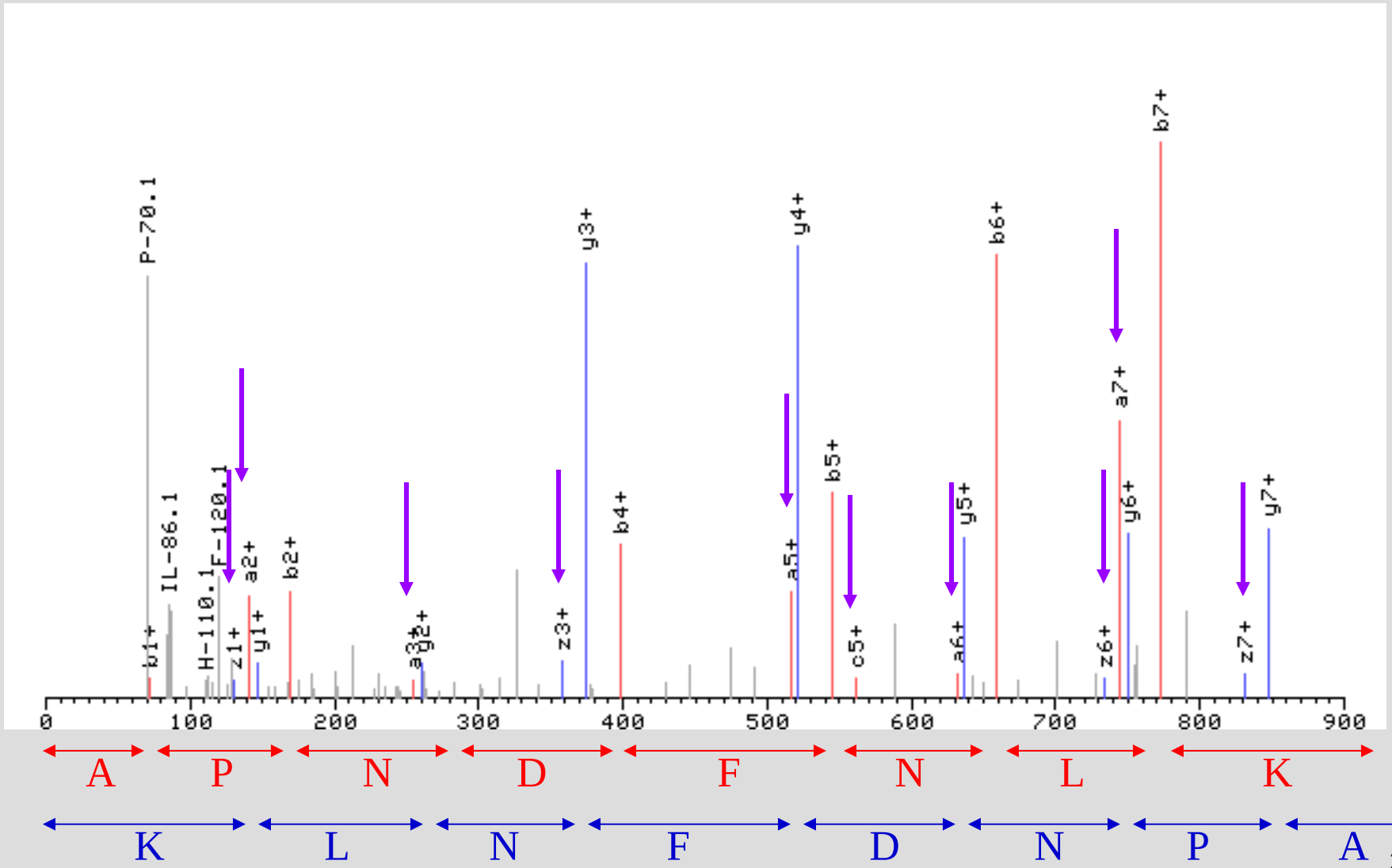
B-ions

Sequence vs. Tandem Mass Spectrum



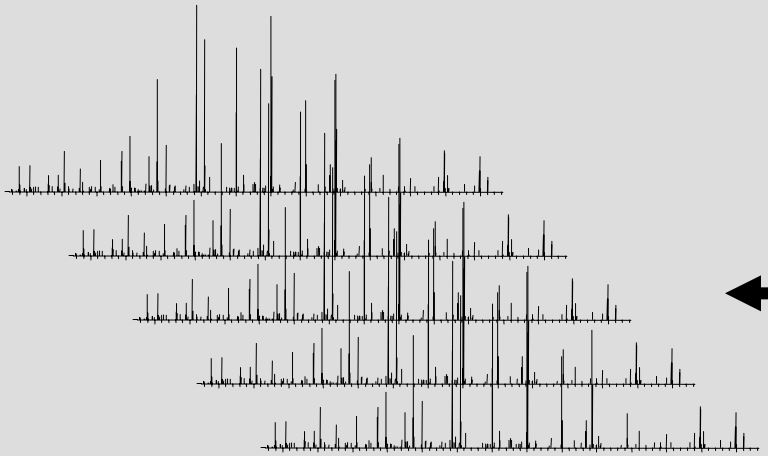
Y-ions

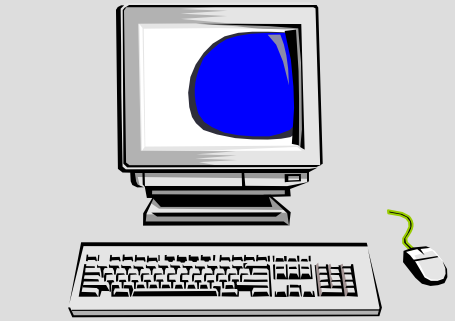
Sequence vs. Tandem Mass Spectrum



Uninterpreted MS/MS Database Search

Raw, uninterpreted
MS/MS spectra





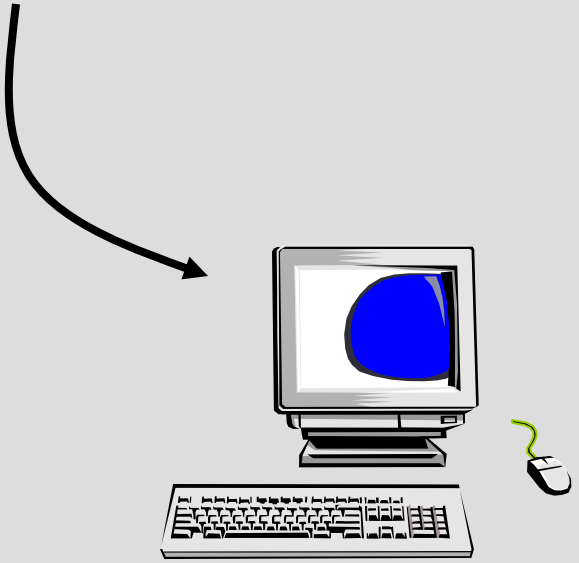
Sequence Database

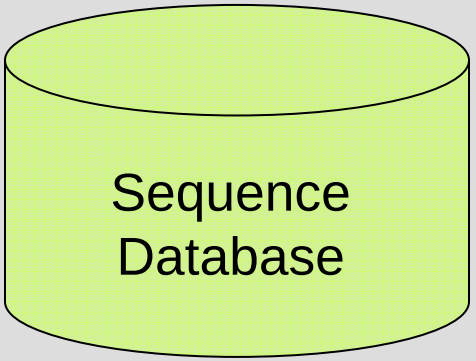
>SEQ1
CVVEELCPTPEGKDIGES
VDLLKLQWCWENGTLRSL
DCDVVS
>SEQ2
DLRSWTVRIDALNHGVKP
HPPNVSVVDLTNR>

Uninterpreted MS/MS Database Search

Input:

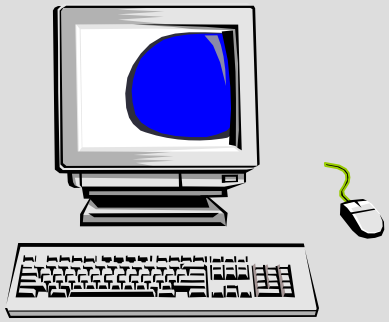
- Fragmentation spectrum
- Precursor mass, charge state





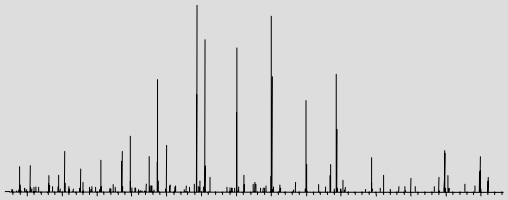
1. From database, select peptides that equal the input mass
2. Theoretically fragment peptides
3. Compare theoretical fragments to acquired spectrum
4. Generate score
5. Rank by score and display best matches

Uninterpreted MS/MS Database Search



Sequence Database

Raw MS/MS spectra



Similarity score

1.00

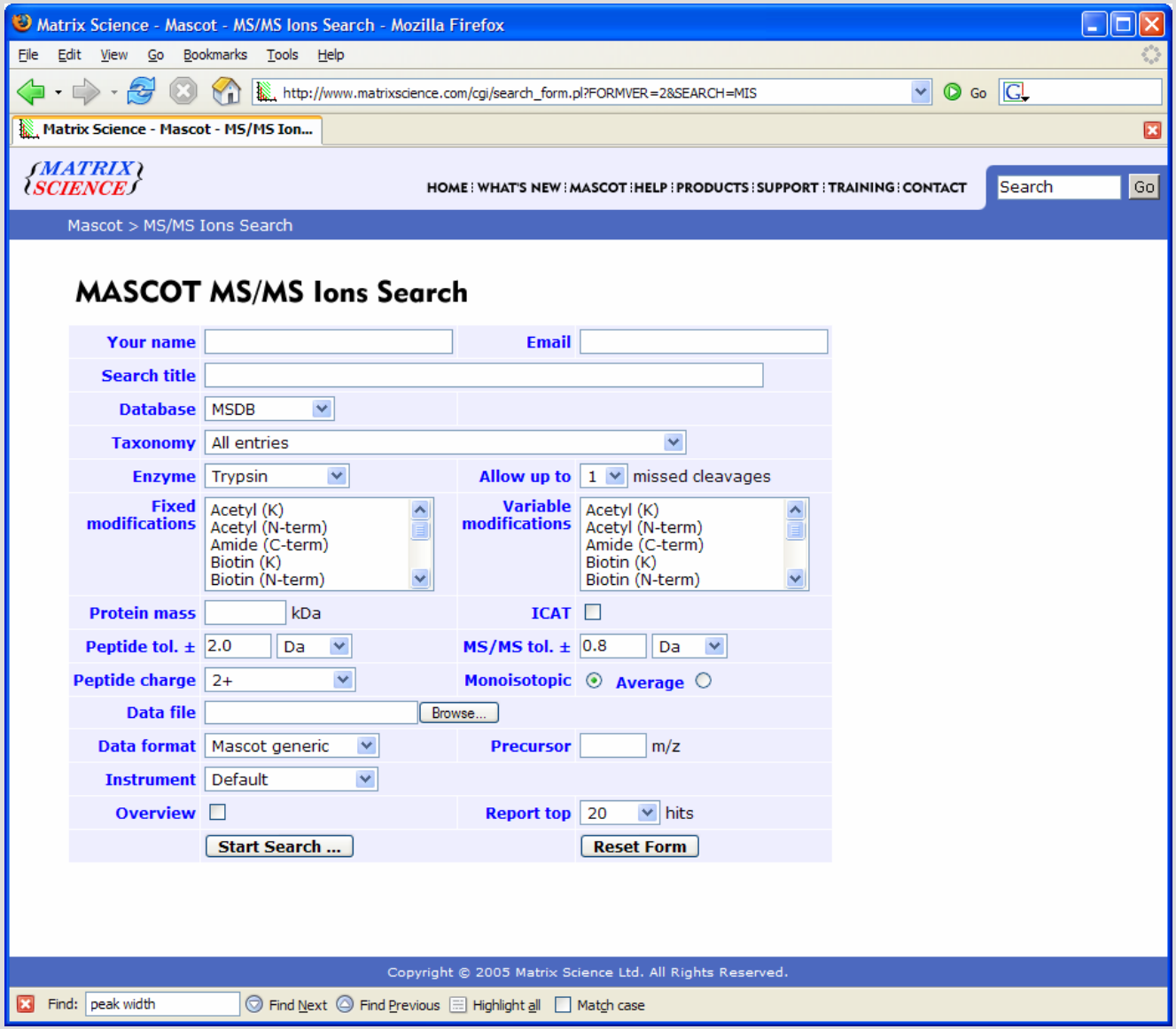
0.34

0.29

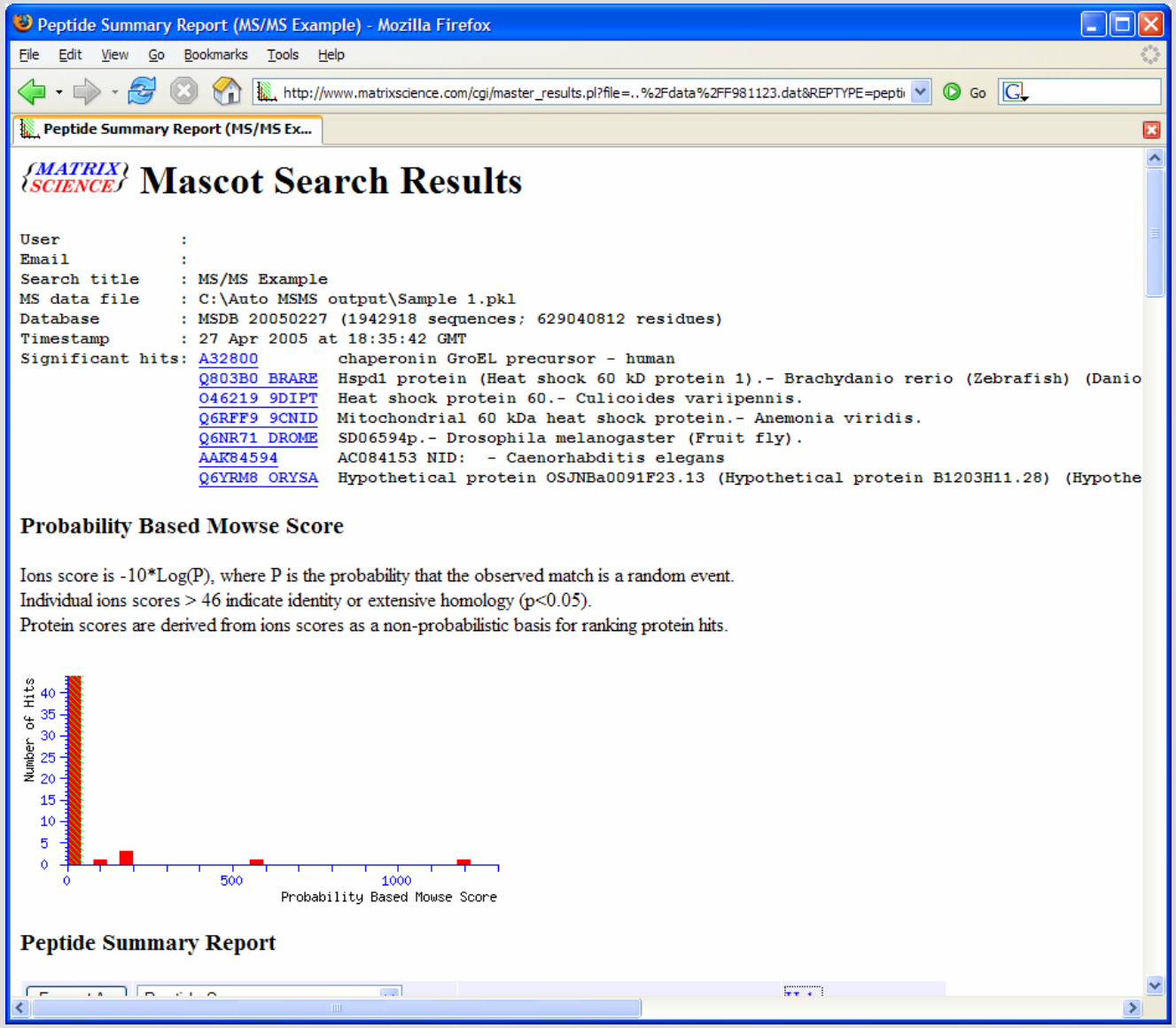
Peptides of same nominal mass

```
>SEQ1
CVVRELCPTPEGKDIGES
VDLLKLQWCWENGTLRSL
DCDVVSRDIGSESTEDRA
MEDIK
>SEQ2
DLRSWTVRIDALNHGVKP
HPPNVSVVDLTNRGDVEK
GKKIFVQKCAQCHTVEKG
GKHKT
```

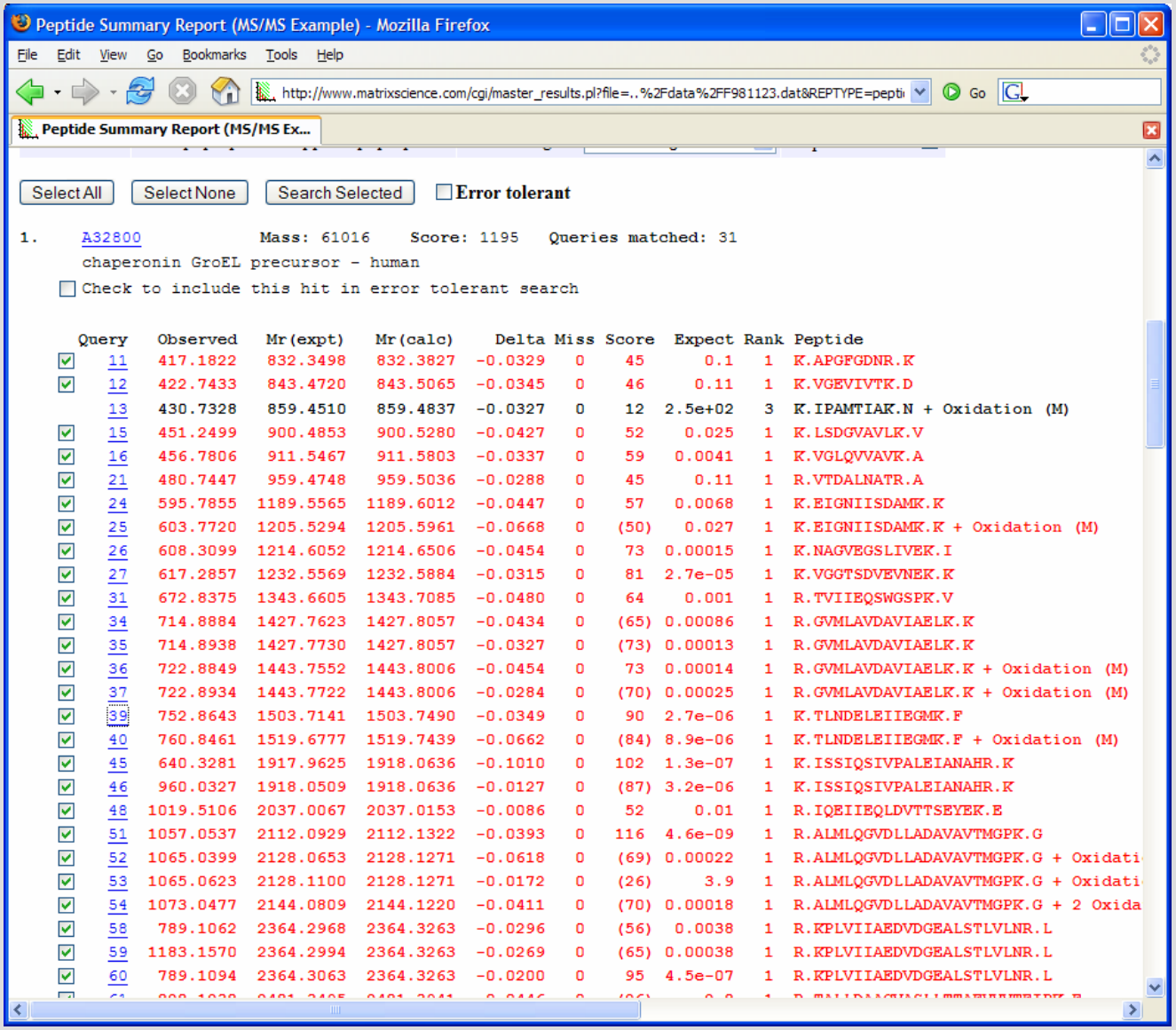
MASCOT


Matrix Science - Mascot - MS/MS Ions Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS
Matrix Science - Mascot - MS/MS Ion...
MATRIX SCIENCE
HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | TRAINING | CONTACT
Search
Go
Mascot > MS/MS Ions Search
MASCOT MS/MS Ions Search
Your name
Email
Search title
Database
MSDB
Taxonomy
All entries
Enzyme
Trypsin
Allow up to
1
missed cleavages
Fixed modifications
Acetyl (K)
Acetyl (N-term)
Amide (C-term)
Biotin (K)
Biotin (N-term)
Variable modifications
Acetyl (K)
Acetyl (N-term)
Amide (C-term)
Biotin (K)
Biotin (N-term)
Protein mass
kDa
ICAT
Peptide tol. ±
2.0
Da
MS/MS tol. ±
0.8
Da
Peptide charge
2+
Monoisotopic
Average
Data file
Browse...
Data format
Mascot generic
Precursor
m/z
Instrument
Default
Overview
Report top
20
hits
Start Search ...
Reset Form
Copyright © 2005 Matrix Science Ltd. All Rights Reserved.
Find: peak width
Find Next
Find Previous
Highlight all
Match case

MASCOT

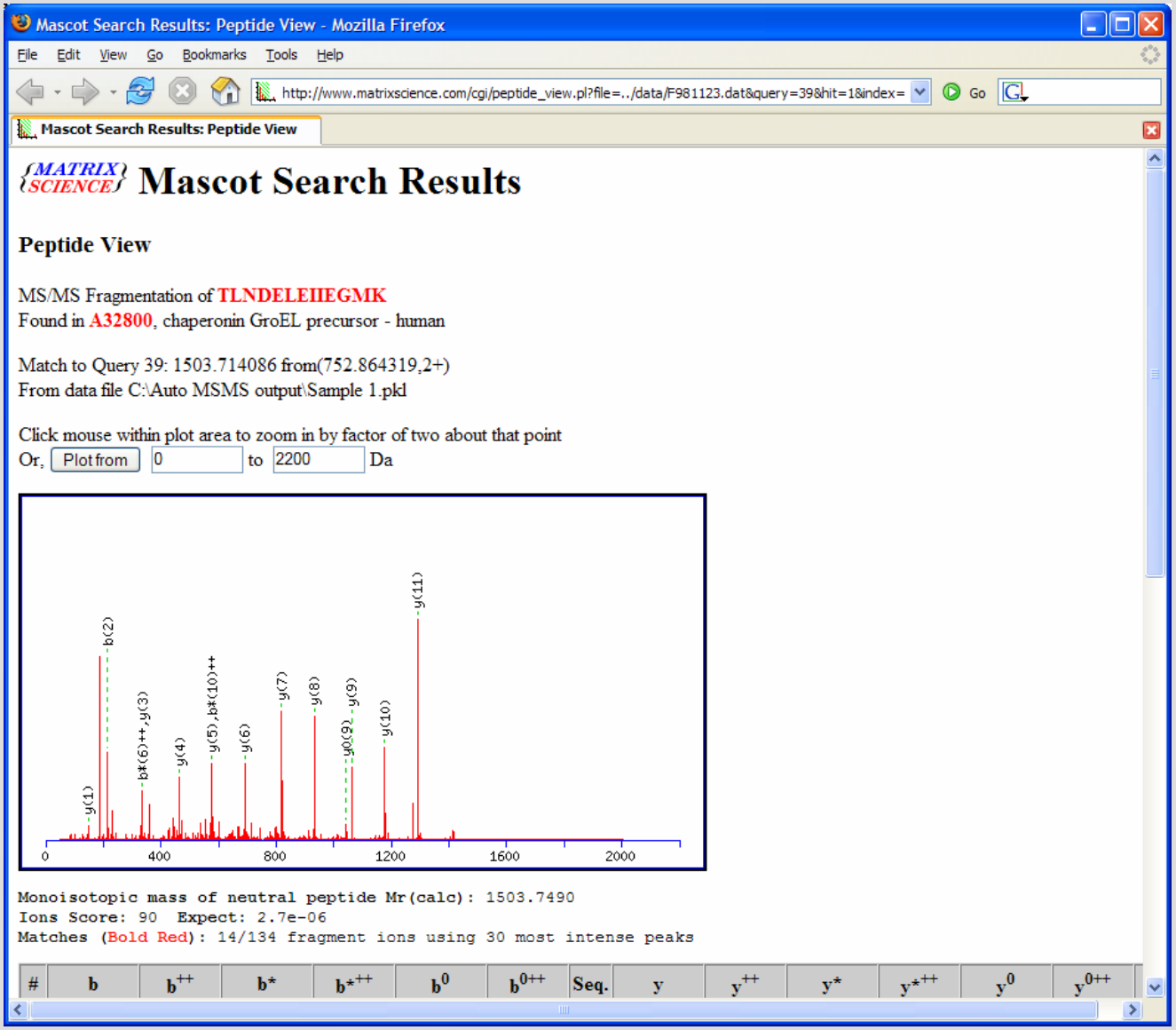

Peptide Summary Report (MS/MS Example) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.matrixscience.com/cgi/master_results.pl?file=..%2Fdata%2FF981123.dat&REPTYPE=pepti
Go
Peptide Summary Report (MS/MS Ex...
MATRIX SCIENCE
Mascot Search Results
User :
Email :
Search title : MS/MS Example
MS data file : C:\Auto MSMS output\Sample 1.pkl
Database : MSDB 20050227 (1942918 sequences; 629040812 residues)
Timestamp : 27 Apr 2005 at 18:35:42 GMT
Significant hits: A32800 chaperonin GroEL precursor - human
Q803B0_BRARE Hspd1 protein (Heat shock 60 kD protein 1).- Brachydanio rerio (Zebrafish) (Danio
O46219_9DIPT Heat shock protein 60.- Culicoides variipennis.
Q6RFF9_9CNID Mitochondrial 60 kDa heat shock protein.- Anemonia viridis.
Q6NR71_DROME SD06594p.- Drosophila melanogaster (Fruit fly).
AAK84594 AC084153 NID: - Caenorhabditis elegans
Q6YRM8_ORYSA Hypothetical protein OSJNBa0091F23.13 (Hypothetical protein B1203H11.28) (Hypothe
Probability Based Mowse Score
Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 46 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.
Number of Hits
0
5
10
15
20
25
30
35
40
0
500
1000
Probability Based Mowse Score
Peptide Summary Report

MASCOT



Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
11	417.1822	832.3498	832.3827	-0.0329	0	45	0.1	1	K.APGFGDNR.K
12	422.7433	843.4720	843.5065	-0.0345	0	46	0.11	1	K.VGEVIVTK.D
13	430.7328	859.4510	859.4837	-0.0327	0	12	2.5e+02	3	K.IPAMTIAK.N + Oxidation (M)
15	451.2499	900.4853	900.5280	-0.0427	0	52	0.025	1	K.LSDGVAVLK.V
16	456.7806	911.5467	911.5803	-0.0337	0	59	0.0041	1	K.VGLQVVAVK.A
21	480.7447	959.4748	959.5036	-0.0288	0	45	0.11	1	R.VTDALNATR.A
24	595.7855	1189.5565	1189.6012	-0.0447	0	57	0.0068	1	K.EIGNIISDAMK.K
25	603.7720	1205.5294	1205.5961	-0.0668	0	(50)	0.027	1	K.EIGNIISDAMK.K + Oxidation (M)
26	608.3099	1214.6052	1214.6506	-0.0454	0	73	0.00015	1	K.NAGVEGSLIVEK.I
27	617.2857	1232.5569	1232.5884	-0.0315	0	81	2.7e-05	1	K.VGGTSDVEVNEK.K
31	672.8375	1343.6605	1343.7085	-0.0480	0	64	0.001	1	R.TVIIEQSWGSPK.V
34	714.8884	1427.7623	1427.8057	-0.0434	0	(65)	0.00086	1	R.GVMLAVDAVIAELK.K
35	714.8938	1427.7730	1427.8057	-0.0327	0	(73)	0.00013	1	R.GVMLAVDAVIAELK.K
36	722.8849	1443.7552	1443.8006	-0.0454	0	73	0.00014	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
37	722.8934	1443.7722	1443.8006	-0.0284	0	(70)	0.00025	1	R.GVMLAVDAVIAELK.K + Oxidation (M)
39	752.8643	1503.7141	1503.7490	-0.0349	0	90	2.7e-06	1	K.TLNDELEIIEGMK.F
40	760.8461	1519.6777	1519.7439	-0.0662	0	(84)	8.9e-06	1	K.TLNDELEIIEGMK.F + Oxidation (M)
45	640.3281	1917.9625	1918.0636	-0.1010	0	102	1.3e-07	1	K.ISSIQSIVPALEIANAHR.K
46	960.0327	1918.0509	1918.0636	-0.0127	0	(87)	3.2e-06	1	K.ISSIQSIVPALEIANAHR.K
48	1019.5106	2037.0067	2037.0153	-0.0086	0	52	0.01	1	R.IQEIIEQLDVTTSEYEK.E
51	1057.0537	2112.0929	2112.1322	-0.0393	0	116	4.6e-09	1	R.ALMLQGVDLLADAVAVTMGPK.G
52	1065.0399	2128.0653	2128.1271	-0.0618	0	(69)	0.00022	1	R.ALMLQGVDLLADAVAVTMGPK.G + Oxidati
53	1065.0623	2128.1100	2128.1271	-0.0172	0	(26)	3.9	1	R.ALMLQGVDLLADAVAVTMGPK.G + Oxidati
54	1073.0477	2144.0809	2144.1220	-0.0411	0	(70)	0.00018	1	R.ALMLQGVDLLADAVAVTMGPK.G + 2 Oxida
58	789.1062	2364.2968	2364.3263	-0.0296	0	(56)	0.0038	1	R.KPLVIIAEDVDGEALSTLVLNR.L
59	1183.1570	2364.2994	2364.3263	-0.0269	0	(65)	0.00038	1	R.KPLVIIAEDVDGEALSTLVLNR.L
60	789.1094	2364.3063	2364.3263	-0.0200	0	95	4.5e-07	1	R.KPLVIIAEDVDGEALSTLVLNR.L

MASCOT


Mascot Search Results: Peptide View - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.matrixscience.com/cgi/peptide_view.pl?file=../data/F981123.dat&query=39&hit=1&index=
Go
Mascot Search Results: Peptide View
MATRIX SCIENCE
Mascot Search Results
Peptide View
MS/MS Fragmentation of TLNDELEIIEGMK
Found in A32800, chaperonin GroEL precursor - human
Match to Query 39: 1503.714086 from(752.864319,2+)
From data file C:\Auto MSMS output\Sample 1.pkl
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 0 to 2200 Da
y(1)
b(2)
b*(6)++,y(3)
y(4)
y(5),b*(10)++
y(6)
y(7)
y(8)
y(9)
y0(9)
y(10)
y(11)
0
400
800
1200
1600
2000
Monoisotopic mass of neutral peptide Mr(calc): 1503.7490
Ions Score: 90 Expect: 2.7e-06
Matches (Bold Red): 14/134 fragment ions using 30 most intense peaks
b b++ b* b*++ b0 b0++ Seq. y y++ y* y*++ y0 y0++

Mascot Score?

From presentation on MatrixScience web site:

- Each ion series is matched and scored independently
- If an ion series contains only a random number of matches, or less, it is discarded
- All combinations of the ion series with non-random levels of matching are tested to see which combination will give the highest score
- Having “too many” ion series doesn’t affect the score, it just reduces specificity

Interpreting Mascot results

- Ions Score = -10 x Log(P)
  - Calculation of P is ‘black box’
  - Extension of the MOWSE score

Interpreting Mascot results

- Identity threshold = -10 x Log(E/N)
  - E is the significance threshold
  - N is the number of peptides in the database matching the precursor mass
- Example
  - If you can accept a 1 in 20 chance of a false positive select an E of 0.05
  - If there are 4000 peptides that match the precursor ion mass

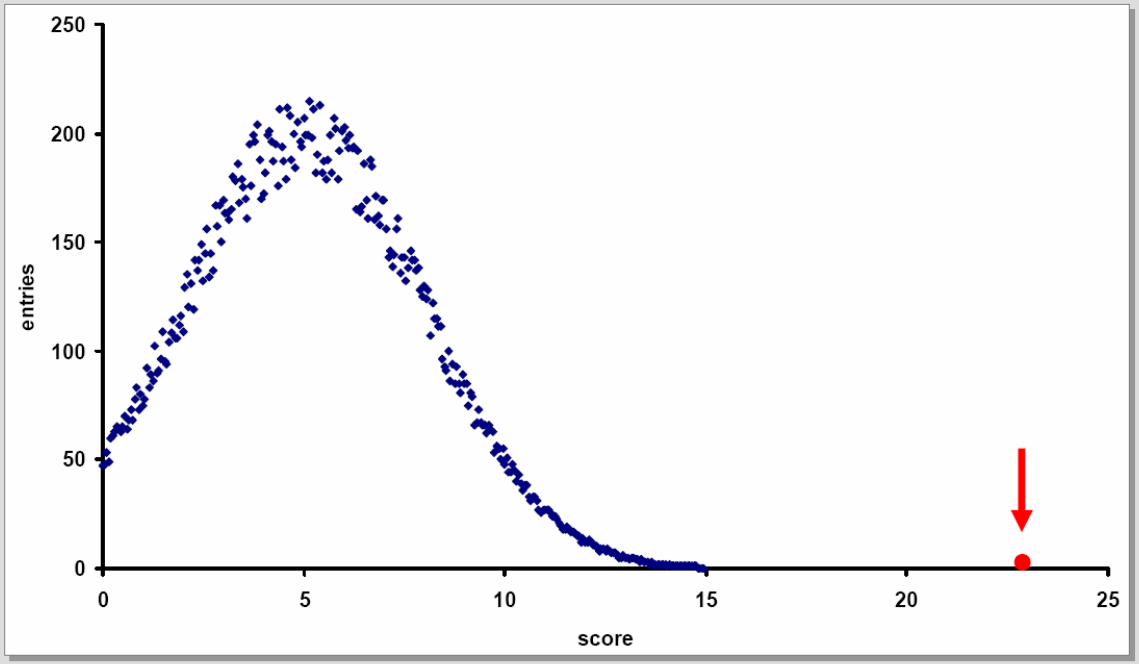
    S = -10 x Log(0.05/4000)

    = 49

Matrix Science http://www.matrixscience.com/pdf/2005WKSHP4.pdf

Interpreting Mascot results

- Homology threshold
  - "The homology threshold is an empirical measure of whether the match is an outlier"



Matrix Science http://www.matrixscience.com/pdf/2005WKSHP4.pdf

Interpreting Mascot results

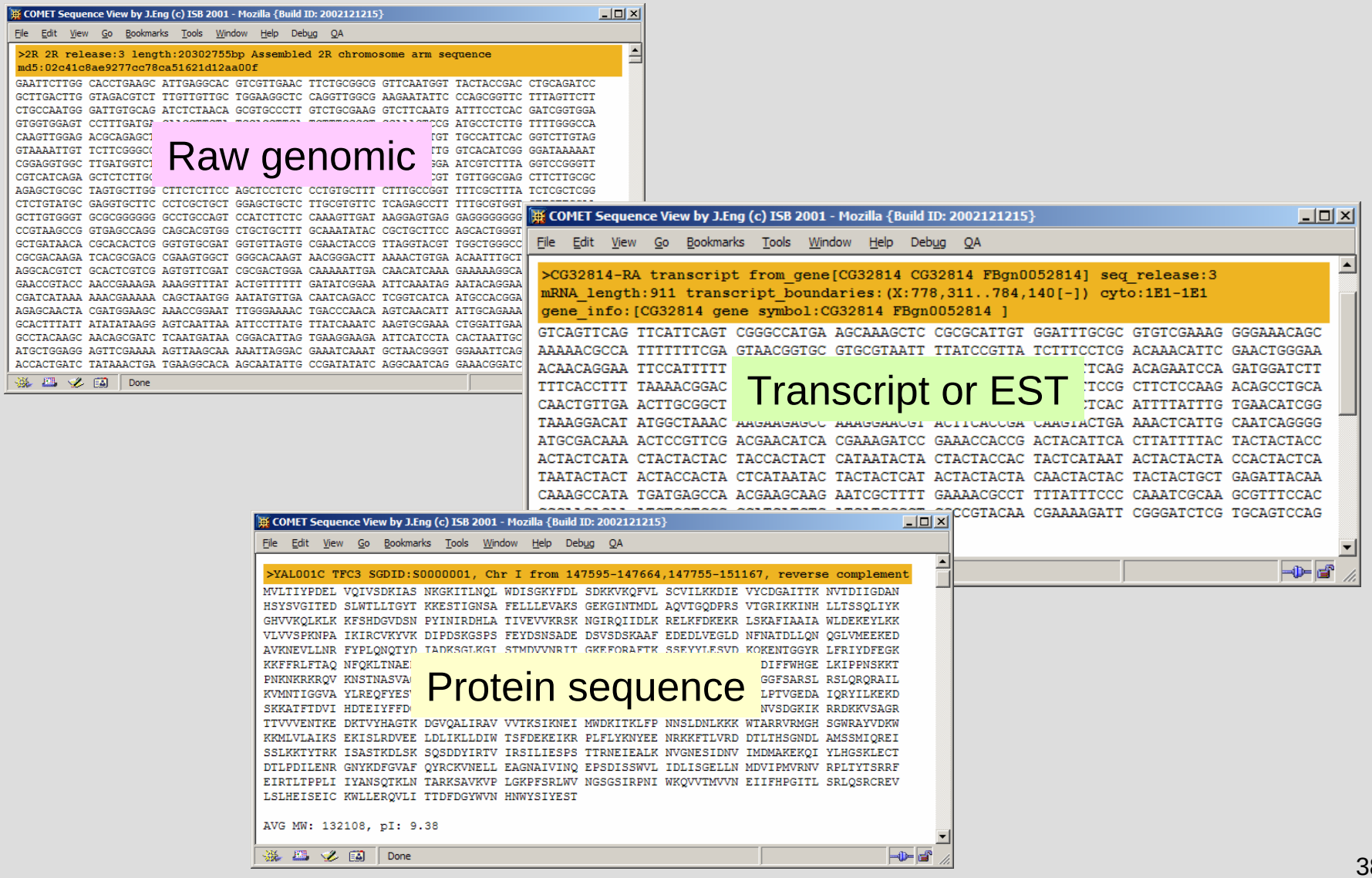
- Expectation value
  - The number of times you could expect to get this score or better by chance
- $E = P_{thresh} \times (10 \wedge ((S_{thresh} - score) / 10))$
- If $P_{thresh} = 0.05$ and $S_{thresh} = 50$
  - score = 40 corresponds to E = 0.5
  - score = 50 corresponds to E = 0.05
  - score = 60 corresponds to E = 0.005

Matrix Science http://www.matrixscience.com/pdf/2005WKSHP4.pdf

MS/MS Database Search Parameters

- Protein, nucleic acid, and EST sequence databases
- Optionally include enzyme specificity in the search
- Post-translation modifications can be identified
- Search software

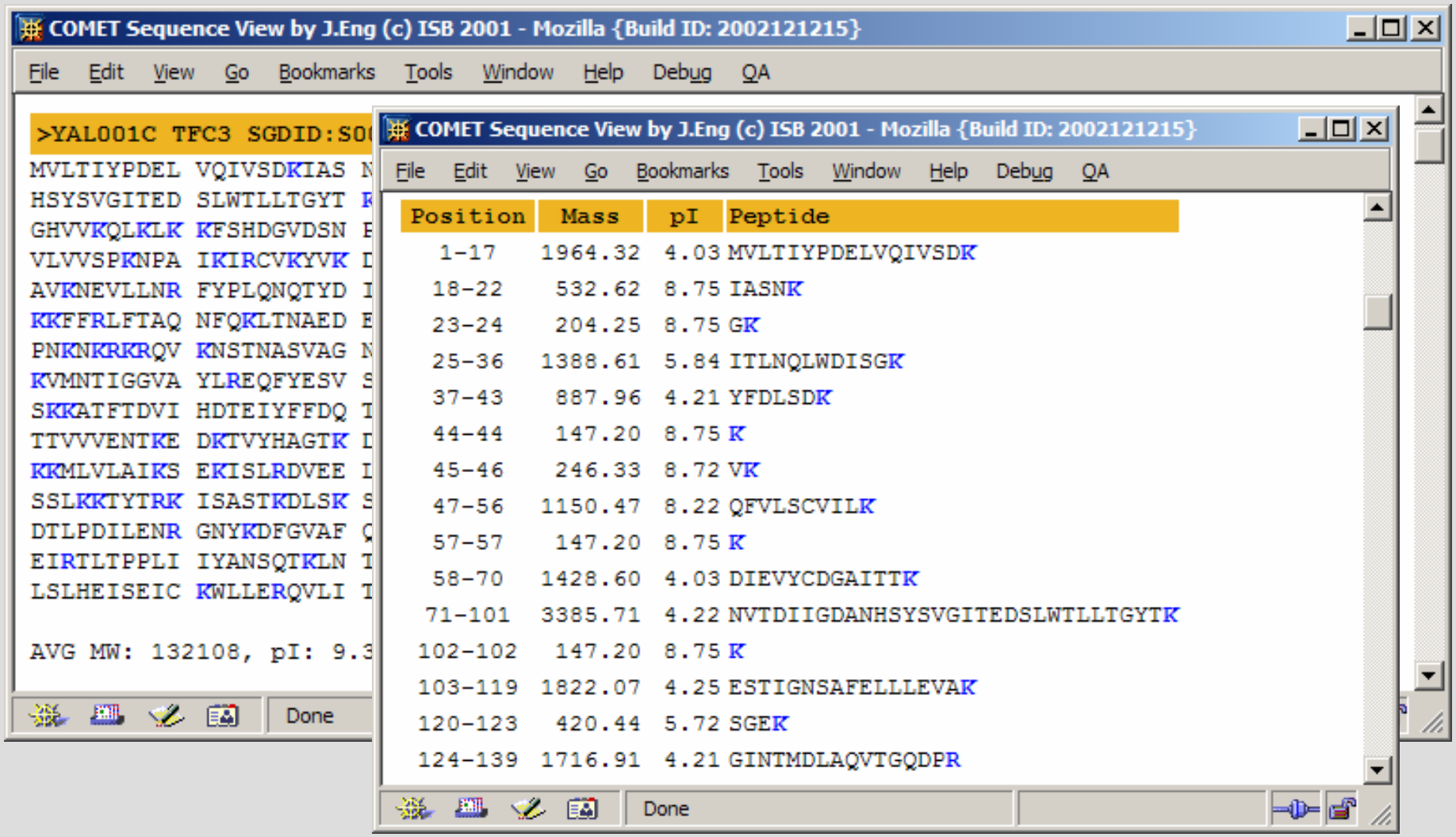
Sequence Databases



MS/MS Database Search Parameters

- Protein, nucleic acid, and short EST sequence databases can all be searched
- Optionally include enzyme specificity in the search
- Post-translation modifications can be identified
- Search software

DB: enzyme constraint



DB: enzyme constraint

tryptic peptides:

GDVEKGTKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGSK

TGQAPGFSYTDANKNKGITWGEETLMEYLENPKSYIPGT

enzyme-unconstrained peptides:

GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRK

TGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGT

DB: tryptic peptides vs. unconstrained search

mass	# tryptic peptides	# unconstr. peptides	factor
1000 Da	1,430	321,999	225x
2000 Da	466	325,096	697x
3000 Da	249	317,750	1276x

human IPI database, 47,754

MS/MS Database Search Parameters

- Protein, nucleic acid, and short EST sequence databases can all be searched
- Optionally include enzyme specificity in the search
- Post-translation modifications can be identified
- Search software

Post-Translation Modifications

- Static Modification
  - ***All*** occurrences of an amino acid is modified
- Variable/Differential Modification
  - One or more occurrences of an amino acid ***may*** be modified
- Modifications can typically be specified on any residue(s) or termini.

Variable Modifications

Serine phosphorylation:

1. DIGSESTEDQAMEDYK
2. DIGSESTEDQAMEDYK
3. DIGSESTEDQAMEDYK
4. DIGSESTEDQAMEDYK

How many peptide forms are possible if you consider serine and threonine phosphorylation for the above peptide? Serine + threonine + tyrosine?

Variable Modification Search

mass	# tryptic peptides	phos STY tryptic	factor	unconstr phos STY
1000 Da	1,430	5,093	3.5x	1,167,740
2000 Da	466	7,283	15.6x	4,538,383
3000 Da	249	16,761	67.3x	15,641,722

human IPI database, 47,754

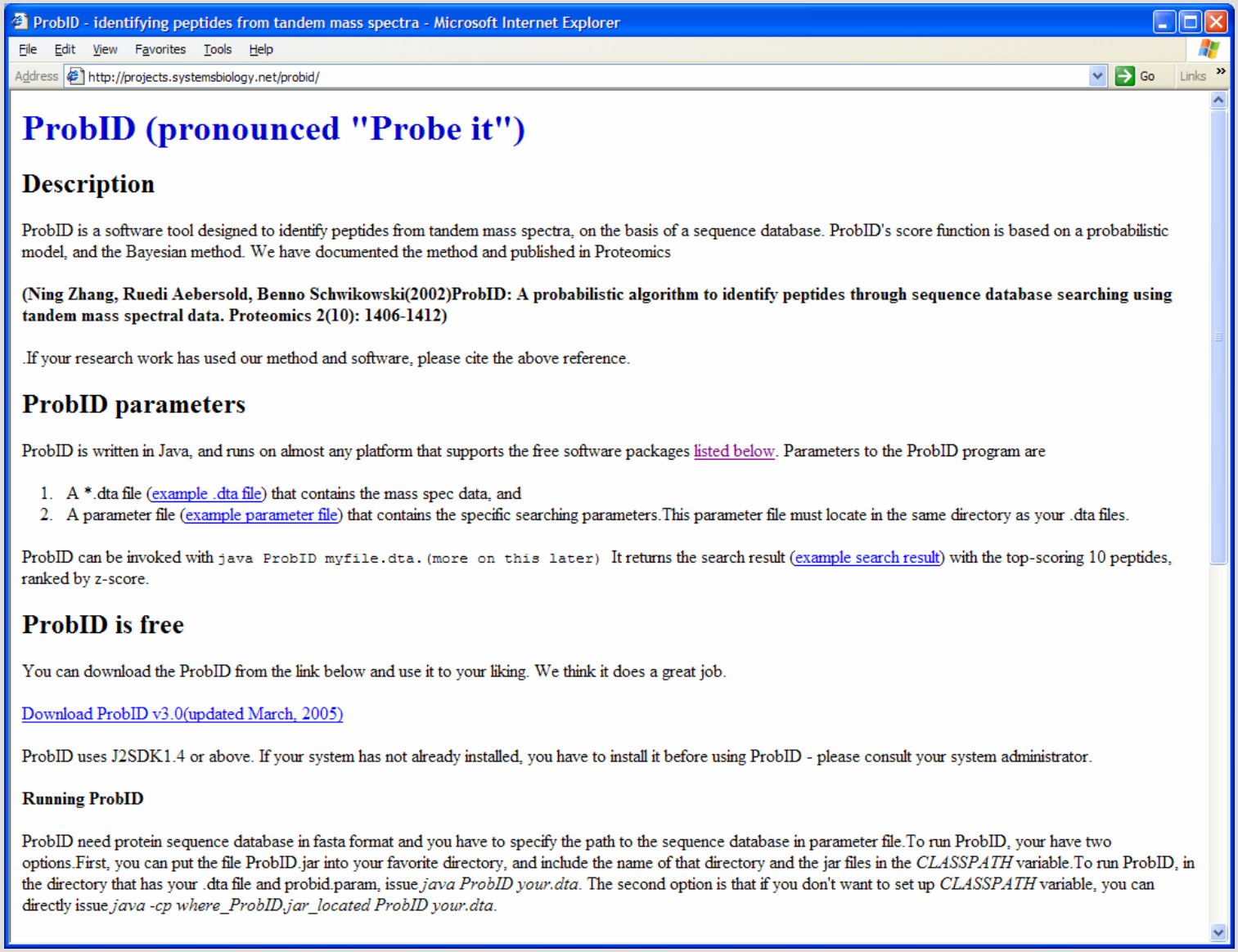
Uninterpreted MS/MS Database Search

- Protein, nucleic acid, and short EST sequence databases can all be searched
- Optionally include enzyme specificity in the search
- Post-translation modifications can be identified
- Search software

What about other programs?

- Phenyx
- SpectrumMill
- ProteinPilot
- SEQUEST
- X! Tandem
- OMSSA
- ProbID

ProbID


ProbID - identifying peptides from tandem mass spectra - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://projects.systemsbiology.net/probid/
Go
Links
ProbID (pronounced "Probe it")
Description
ProbID is a software tool designed to identify peptides from tandem mass spectra, on the basis of a sequence database. ProbID's score function is based on a probabilistic model, and the Bayesian method. We have documented the method and published in Proteomics
(Ning Zhang, Ruedi Aebersold, Benno Schwikowski(2002)ProbID: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. Proteomics 2(10): 1406-1412)
.If your research work has used our method and software, please cite the above reference.
ProbID parameters
ProbID is written in Java, and runs on almost any platform that supports the free software packages listed below. Parameters to the ProbID program are
1. A *.dta file (example .dta file) that contains the mass spec data, and
2. A parameter file (example parameter file) that contains the specific searching parameters.This parameter file must locate in the same directory as your .dta files.
ProbID can be invoked with java ProbID myfile.dta.(more on this later) It returns the search result (example search result) with the top-scoring 10 peptides, ranked by z-score.
ProbID is free
You can download the ProbID from the link below and use it to your liking. We think it does a great job.
Download ProbID v3.0(updated March, 2005)
ProbID uses J2SDK1.4 or above. If your system has not already installed, you have to install it before using ProbID - please consult your system administrator.
Running ProbID
ProbID need protein sequence database in fasta format and you have to specify the path to the sequence database in parameter file.To run ProbID, your have two options.First, you can put the file ProbID.jar into your favorite directory, and include the name of that directory and the jar files in the CLASSPATH variable.To run ProbID, in the directory that has your .dta file and probid.param, issue java ProbID your.dta. The second option is that if you don't want to set up CLASSPATH variable, you can directly issue java -cp where_ProbID.jar_located ProbID your.dta.

ProbID

The Bayesian approach that is the basis of the score function attempts to quantify our belief that the experimental spectrum under consideration was generated from a given candidate peptide sequence k in the database. By S we denote the experimental spectrum; B stands for any other prior background information about the experiment. The Bayesian method specifies the posterior probability $\Pr(k|S, B)$ for a hypothetical peptide k, given S and B, as

$$\Pr(k|S,B) = \frac{\Pr(k|B)^* \Pr(S|k,B)}{\sum_{k_i \in \text{database}} \Pr(k_i|B)^* \Pr(S|k_i,B)} \quad (1)$$

The peptide k with the highest value of $\Pr(k|S, B)$ is the most likely one to have generated the spectrum S. Figure 2 illustrates how $\Pr(k|S, B)$ is composed. The probabilities of each contributing factor are multiplied to yield the final posterior probability. The contributing factors are detailed below in Section 3.2.

ProbID

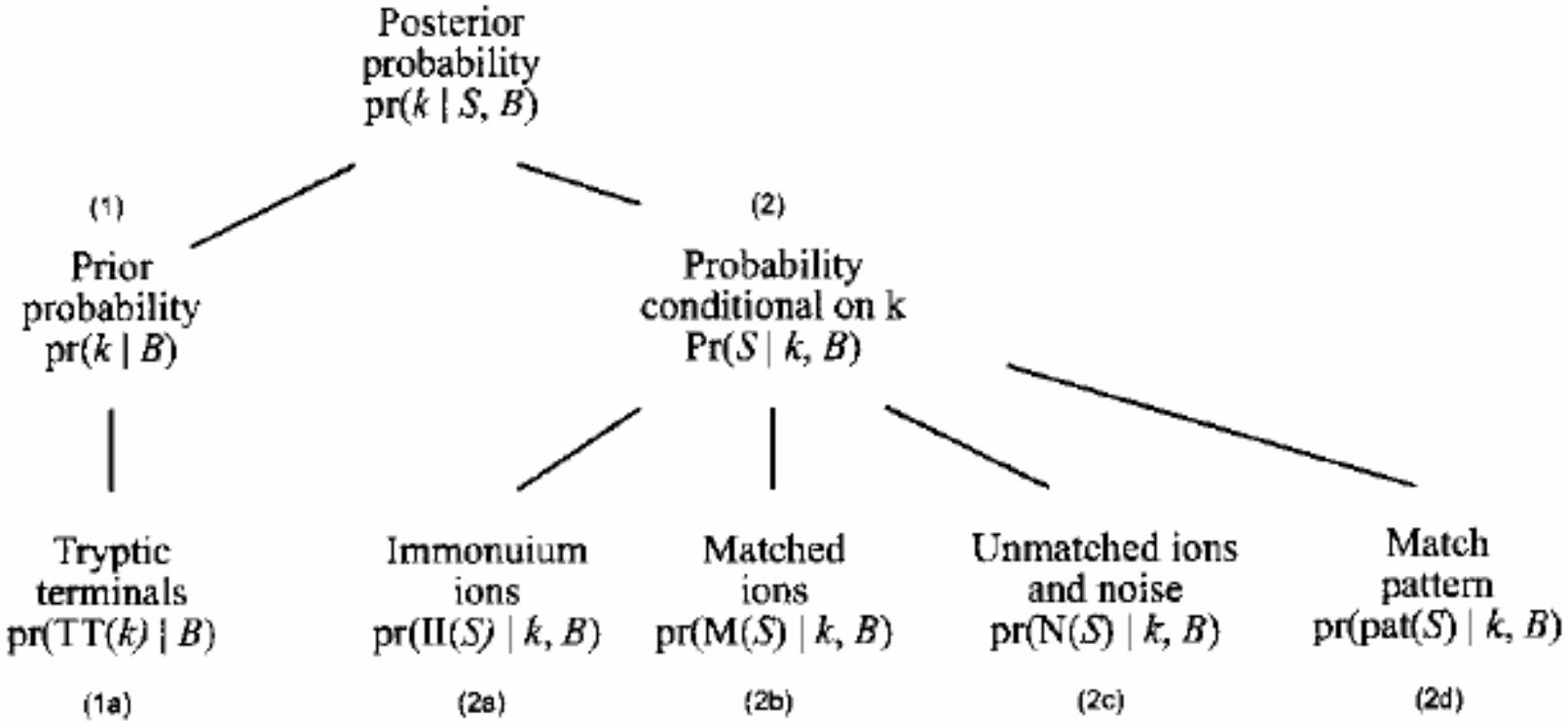


Figure 2. Composition of the ProbID posterior probability function Pr(*k*|S,B)

ProbID

Immonium ions:

H, M, W, Y, F

$pr(II(S)|k,B) = 1 - i/5$

where i = # of immonium peaks in spectrum w/o corresponding amino acid

Unmatched ions:

$pr(N(S)|k,B) = (1/mass_{max} - mass_{min})^r$

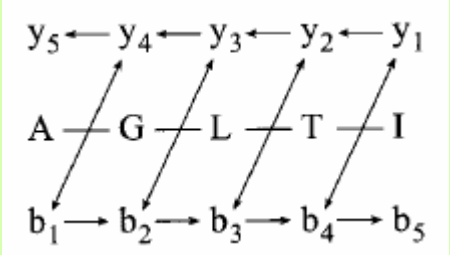
where r = # of unmatched ions and $mass_{max}$ & $mass_{min}$ are the highest and lowest peaks in spectrum

ProbID

Match pattern:

pr(pat(S)|k,B) = (# of matched ion pairs) / 3(n-1)

n = # of AA in peptide



Matched ions:

$$\mathrm{pr}(\mathrm{M}(S)\,|\,k,B) = \prod a_i e^{\frac{(m_i - m)^2}{2\sigma^2}}$$

a_i = amplitude of each peak

m_i = mass of each peak

σ = mass accuracy std dev

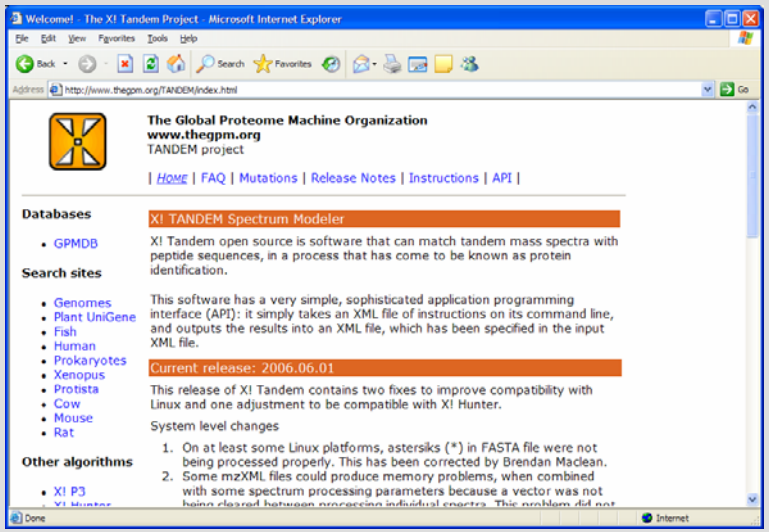
ProbID output

Rank	Peptide Seq.	Poster. prob.	Ions	(M+H)+	Reference
1	**K.VKEAMAPK**	**0.9999**	**13/16**	**874.06323**	**sp|P02666| CASB_BOVIN**
2	K.KVEKELQ	7.9913E-12	11/14	874.00195	HI0814
3	R.QDLEQLQ	3.0763E-15	10/14	873.91504	HI0078
4	**L.KNAISDPK**	**1.0963E-15**	**13/16**	**872.97375**	**HI0329**
5	V.EVKRVDK	2.5961E-16	10/14	874 005	HI0453
6	R.VKQKLEK	6.1706E-17	10/14	873.06055	sp|P04460|
7	V.KNLEEIK	4.5430E-17	11/14	874.	
8	K.ELAARWK	2.1886E-17	11/14	874.	
9	A.KVKQENK	6.8512E-18	10/14	874.	
10	I.KLNEKIK	2.1526E-18	10/14	873.	

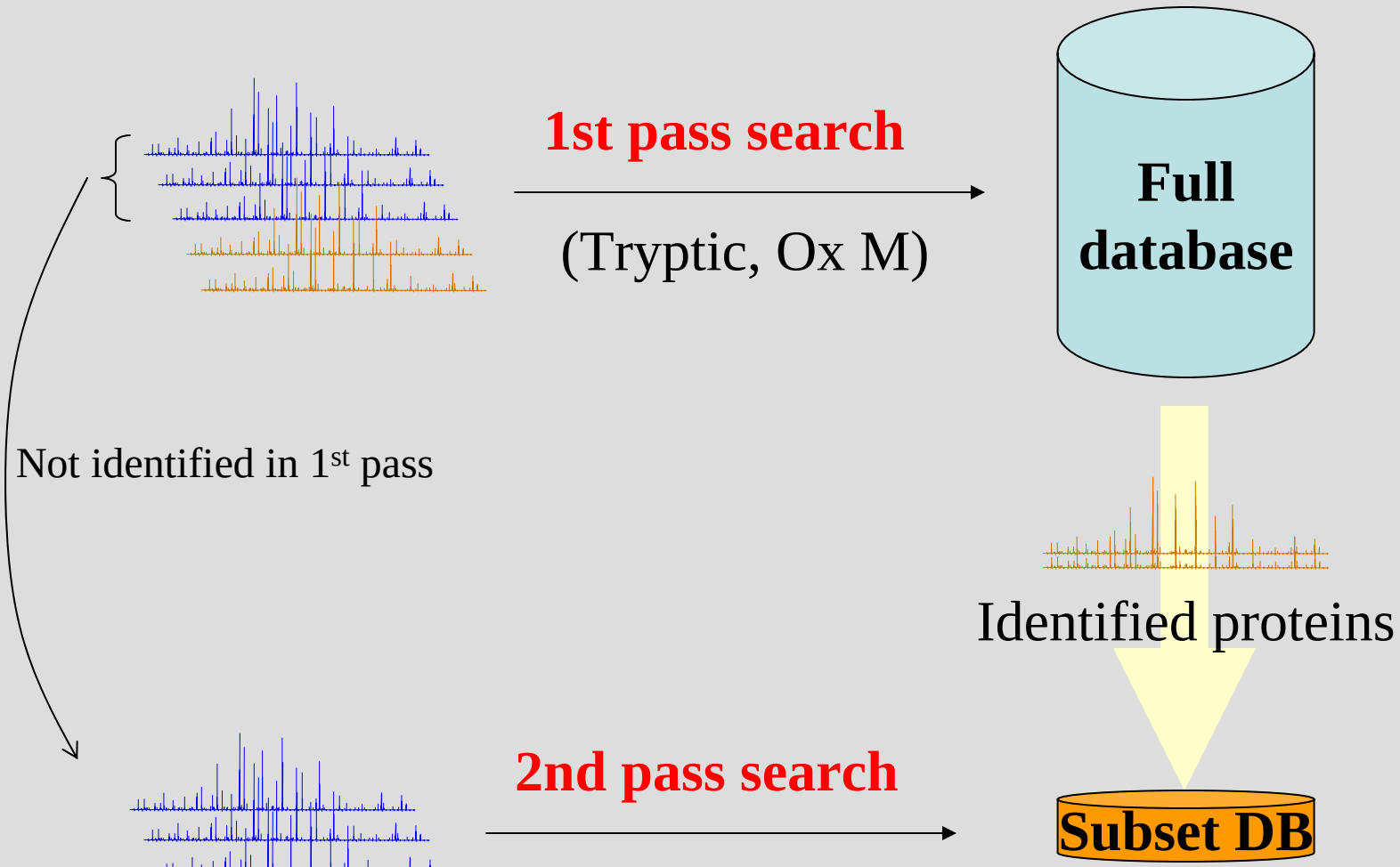
Rank	Peptide Seq.	Poster. prob.	Ions	(M+H)+	Reference
1	R.QQQQKISHLQK	0.3520	18/44	1366.5398	HI0658
2	V.EEVDYEKGRLK	0.3368	18/44	1366.4906	HI0717
3	M.KMLDFLQRGEK	0.3106	20/44	1365.6104	HI0934
4	D.QQFVDECLDRI	5.6733E-4	21/44	1366.5084	HI0724
5	V.VKYVKKLVPKY	5.8052E-6	20/44	1365.7223	HI0390.1
6	R.QNYPDKFANIR	3.0332E-6	17/44	1366.4961	HI1113
7	K.KMLRPFVYRR	1.9713E-7	14/40	1366.6906	HI0069
8	S.EQLHQAWKVVQ	3.2021E-9	19/44	1366.5393	HI1473
9	P.EKSQAFGQKSKK	8.8377E-10	20/48	1366.537	HI1296
10	E.KEKSEFKLELD	2.5863E-10	19/44	1366.531	MYH6_ RABIT

X! Tandem

- Open source search engine
- Very fast
- Lots of user-definable search options
- Built-in “refinement” mode



X! Tandem refinement mode:



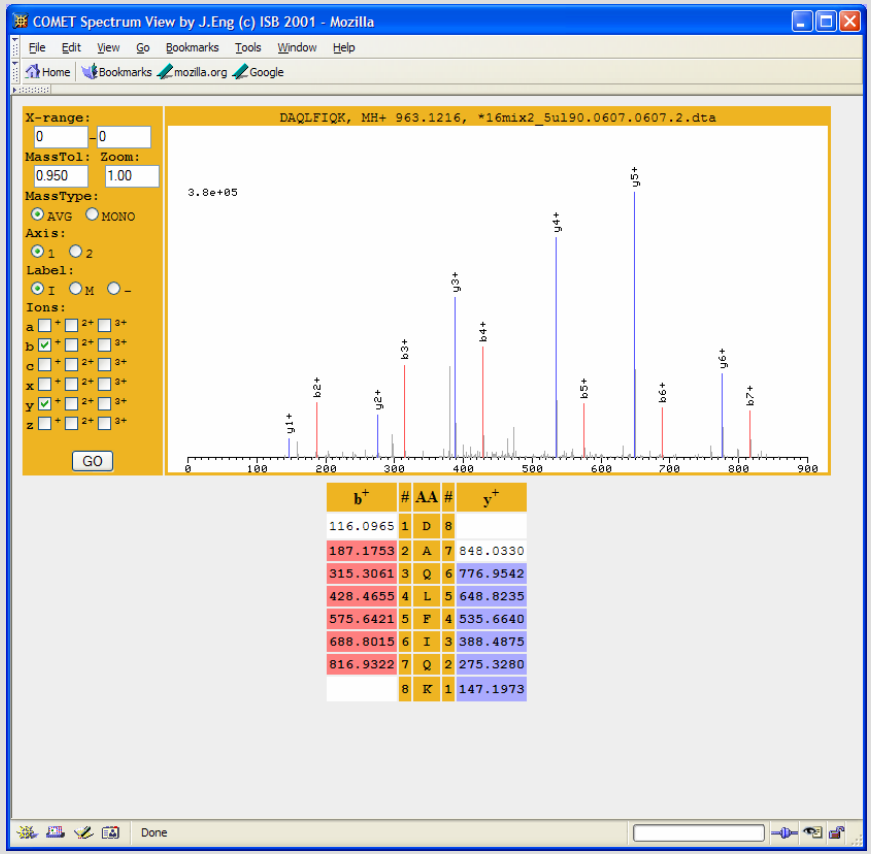
Interpretation Rules

An enzyme un-restricted search can greatly assist in the interpretation process.

Look for peptides that exhibit the expected cleavage at both the N- and C-terminus.

Don’t bother with peptides that exhibit no correct cleavage.

Interpretation Rules



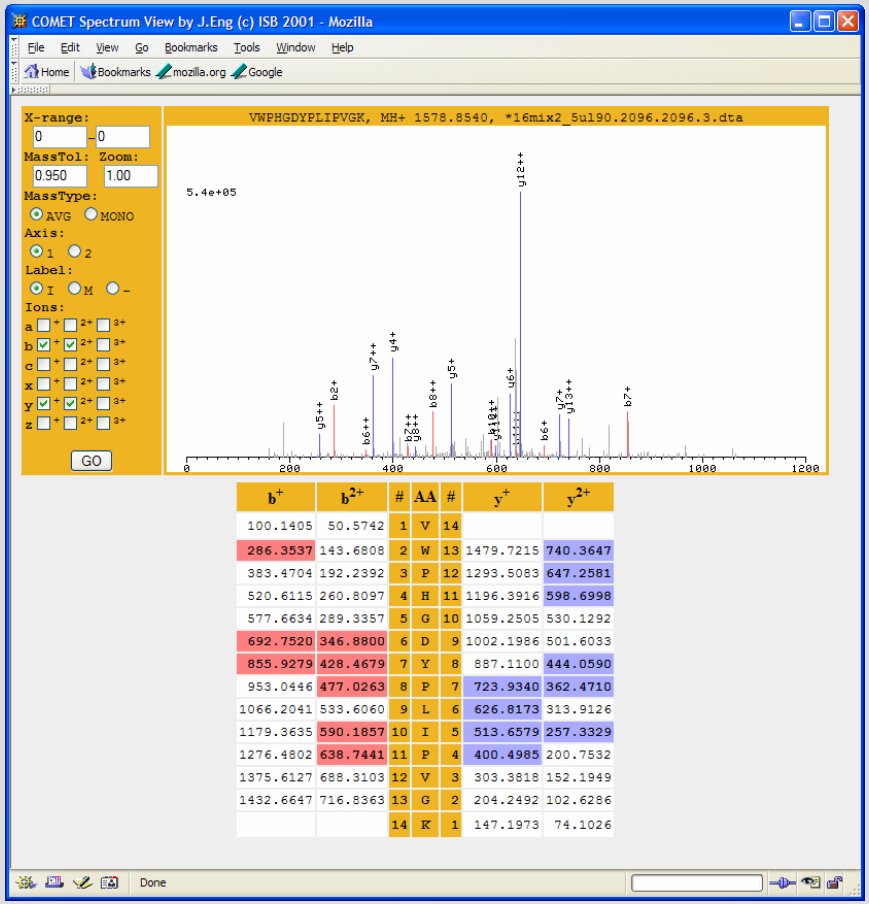
b^+	#	AA	#	y^+
116.0965	1	D	8	
187.1753	2	A	7	848.0330
315.3061	3	Q	6	776.9542
428.4655	4	L	5	648.8235
575.6421	5	F	4	535.6640
688.8015	6	I	3	388.4875
816.9322	7	Q	2	275.3280
	8	K	1	147.1973

Match all fragment ions!

Correct identifications don't exhibit random fragment ion matches. Look for a series of y-ions or b-ions.

Trypsin leaves a basic residue (K or R) at the C-terminus which translate to strong y-ions so hopefully the big peaks match y-ions.

Interpretation Rules



b^+	b^{2+}	#	AA	#	y^+	y^{2+}
100.1405	50.5742	1	V	14		
286.3537	143.6808	2	W	13	1479.7215	740.3647
383.4704	192.2392	3	P	12	1293.5083	647.2581
520.6115	260.8097	4	H	11	1196.3916	598.6998
577.6634	289.3357	5	G	10	1059.2505	530.1292
692.7520	346.8800	6	D	9	1002.1986	501.6033
855.9279	428.4679	7	Y	8	887.1100	444.0590
953.0446	477.0263	8	P	7	723.9340	362.4710
1066.2041	533.6060	9	L	6	626.8173	313.9126
1179.3635	590.1857	10	I	5	513.6579	257.3329
1276.4802	638.7441	11	P	4	400.4985	200.7532
1375.6127	688.3103	12	V	3	303.3818	152.1949
1432.6647	716.8363	13	G	2	204.2492	102.6286
		14	K	1	147.1973	74.1026

If a big peak matches a y-ion from an N-terminal cleavage of proline, that is a good indication of a correct identification.

The reverse is not true: a proline in a peptide that does not correspond to a big peak is not an indication of an incorrect identification.

Random or reverse databases?

Original sequence:

MKWVTFISLLFLFSSAYSRGVFRRDAHKSEVAHR
FKDLGEENFKALVLIAFAQYRQQCPFEDHVKLVN
EVTEFAKTCVADESAENCDKSLHTLFGDKLCTVA

Reverse peptide sequence:

MKSYASSFLFLLSIFTVWRGVFRRHADKHAVESR
FKFNEEGLDKYQAFAILVLARVHDEFPCQQKAFE
TVENVLKDCNEASEDAVCTKDGFLTHLSKAVTCL

When searching forward + reverse sequence database, estimated number of incorrect matches is:

$$\frac{2 * (\text{\# reverse matches passing cutoff})}{\text{\# total matches passing cutoff}}$$