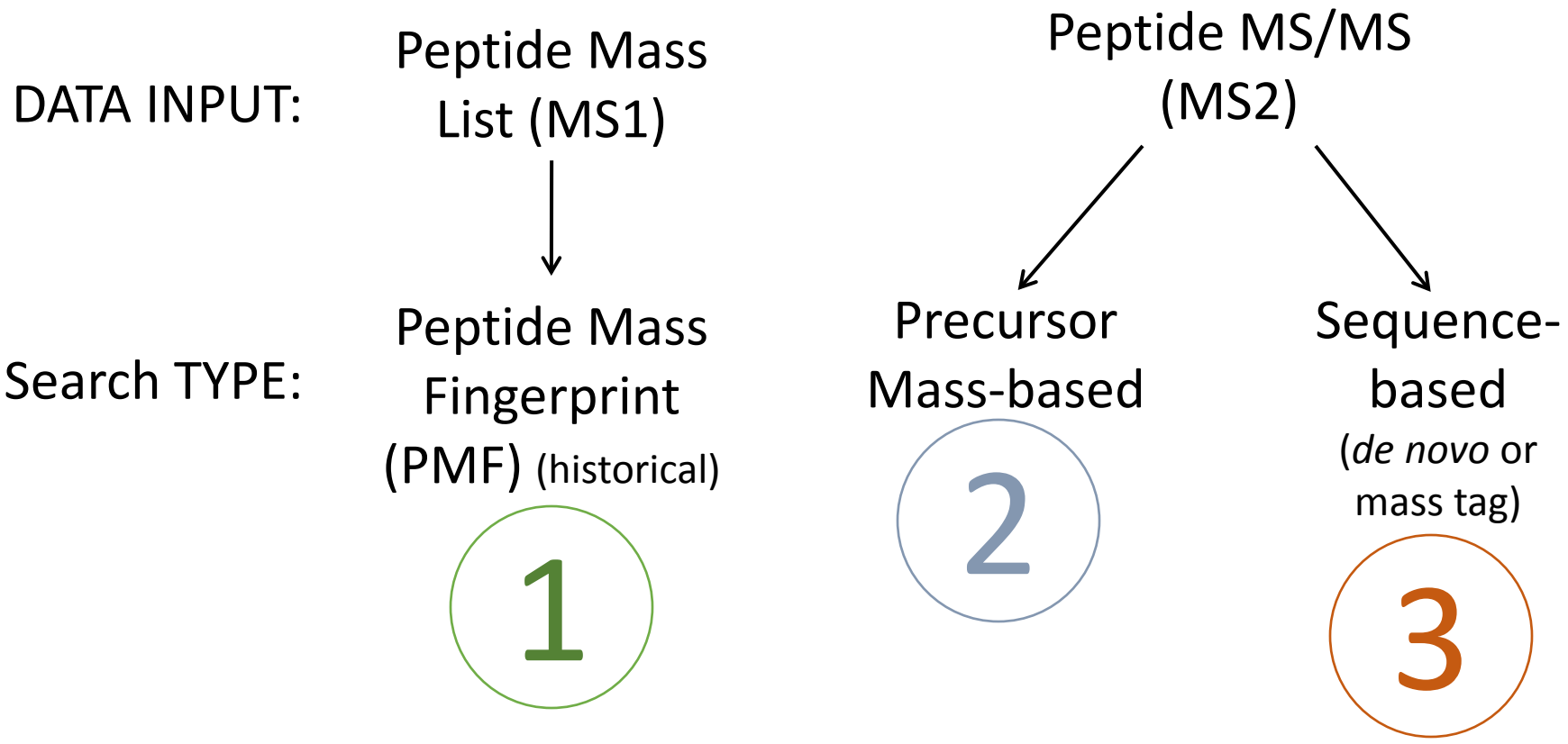
Outline

- Database search types
  - Peptide Mass Fingerprint (PMF)
  - Precursor mass-based
  - Sequence tag
- Results comparison across programs
- Manual inspection of results

Terminology

- Mass tolerance
- MS/MS search
- FASTA
- Tandem MS
- Precursor mass
- Product ion
- Theoretical mass
- Experimental mass
- Sequence tag
- *de novo* sequencing

Three Strategies for Protein Inference from Peptide MS or MS/MS Data



Peptide Mass Fingerprint (PMF)

- Historical: after 1D or 2D in-gel Digestion of single bands or spots
- Current applications: limited; decreasing utility

Component	Description
Input (Data)	Peptide Mass List • Mass type (Mr or MH+) • Average or Monoisotopic
Target	Protein FASTA Sequence Database
Search Parameters	Proteolytic enzyme
	# Missed enzyme cleave sites
	Amino acid modifications
	Mass tolerance

1

Peptide Mass Fingerprint: INPUT

UNKNOWN PROTEIN (*amino acid sequence*):

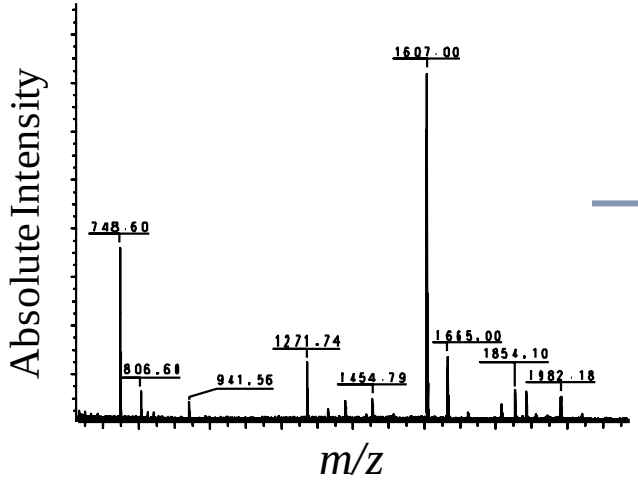
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTG
HPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLT
ALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLE
FISDAIIHVLHSHPGNFGADAQGAMTKALELFRND
IAAKYKELGFQG

A) In gel Trypsin digestion

GLSDGEWQQVLNVWGK
VEADIAGHGQEVLIR
LFTGHPETLEK
TEAEMK
ASEDLK
HGTVVLTALGGILK
GHHEAELKPLAQSHATK
YLEFISDAIIHVLHSHPGNFGADAQGAMTK
ALELFR
NDIAAK
ELGFQG

B) Acquire MALDI-TOF Mass Spectrum of peptide mixture

Desalt using C18 ZipTip; spot on MALDI target



C) Experimental Peptide Mass (or *m/z*) List

m/z	Peptide
631.71	NDIAAK
650.71	ELGFQG
662.72	ASEDLK
708.81	TEAEMK
748.90	ALELFR
1272.44	LFTGHPETLEK
1379.69	HGTVVLTALGGILK
1607.81	VEADIAGHGQEVLIR
1817.01	GLSDGEWQQVLNVWGK
1855.06	GHHEAELKPLAQSHATK
3241.65	YLEFISDAIIHVLHSHPGNFGADAQGAMTK

1 PMF TARGET: Protein FASTA Sequence Database

```
>sp|P31946|1433B_HUMAN 14-3-3 protein beta/alpha OS=Homo sapiens
GN=YWHAB PE=1 SV=3
MTMDKSELVQKAKLAEQAERYDDMAAAMKAVTEQGHELSNEERNLLSVAYKNVVGARRSS
WRVISSIEQKTERNEKKQQMGKEYREKIEAELQDICNDVLELLDKYLIPNATQPESKVFY
LKMKGDYFRYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFY
YEILNSPEKACSLAKTAFDEAIAELDTLNEESYKDSTLIMQLLRDNLTLWTSENQGDEGD
AGEGEN
>sp|P31946-2|1433B_HUMAN Isoform Short of 14-3-3 protein
beta/alpha OS=Homo sapiens GN=YWHAB
MDKSELVQKAKLAEQAERYDDMAAAMKAVTEQGHELSNEERNLLSVAYKNVVGARRSSWR
VISSIEQKTERNEKKQQMGKEYREKIEAELQDICNDVLELLDKYLIPNATQPESKVFYLK
MKGDYFRYLSEVASGDNKQTTVSNSQQAYQEAFEISKKEMQPTHPIRLGLALNFSVFYYE
ILNSPEKACSLAKTAFDEAIAELDTLNEESYKDSTLIMQLLRDNLTLWTSENQGDEGDAG
EGEN
>sp|P62258|1433E_HUMAN 14-3-3 protein epsilon OS=Homo sapiens
GN=YWHAE PE=1 SV=1
MDDREDLVYQAKLAEQAERYDEMVESMKKVAGMDVELTVEERNLLSVAYKNVIGARRASW
RIISSIEQKEENKGGEDKLKMIREYRQMVETELKLICCDILDVLDKHLIPAANTGESKVF
YYKMKGDYHRYLAEFATGNDRKEAAENSLVAYKAASDIAMTELPPTHPIRLGLALNFSVF
YYEILNSPDRACRLAKAAFDDAIAELDTLSEESYKDSTLIMQLLRDNLTLWTSDMQGDGE
EQNKEALQDVEDENQ
>sp|Q04917|1433F_HUMAN 14-3-3 protein eta OS=Homo sapiens GN=YWHAH
PE=1 SV=4
MGDREQLLQRARLAEQAERYDDMASAMKAVTELNEPLSNEDRNLLSVAYKNVVGARRSSW
RVISSIEQKTMADGNEKKLEKVKAYREKIEKELETVCNDVLSLLDKFLIKNCNDFQYESK
VFYLKMKGDYYRYLAEVASGEKKNSVVEASEAAYKEAFEISKEQMQPTHPIRLGLALNFS
VFYYEIQNAPEQACLLAKQAFDDAIAELDTLNEDSYKDSTLIMQLLRDNLTLWTSDQQDE
EAGEGN
>sp|P61981|1433G_HUMAN 14-3-3 protein gamma OS=Homo sapiens
GN=YWHAG PE=1 SV=2
MVDREQLVQKARLAEQAERYDDMAAAMKNVTELNEPLSNEERNLLSVAYKNVVGARRSSW
RVISSIEQKTSADGNEKKIEMVRAYREKIEKELEAVCQDVLSLLDNYLIKNCSETQYESK
VFYLKMKGDYYRYLAEVATGEKRATVVESSEKAYSEAHEISKEHMQPTHPIRLGLALNYS
VFYYEIQNAPEQACHLAKTAFDDAIAELDTLNEDSYKDSTLIMQLLRDNLTLWTSDQQDD
DGGEGNN
>sp|P31947|1433S_HUMAN 14-3-3 protein sigma OS=Homo sapiens GN=SFN
PE=1 SV=1
MERASLIQKAKLAEQAERYEDMAAFMKGAVEKGEELSCEERNLLSVAYKNVVGGQRAAWR
VLSSIEQKSNEEGSEEKGPEVREYREKVETELQGVCDTVLGLLDSHLIKEAGDAESRVFY
LKMKGDYYRYLAEVATGDDKKRIIDSARSAYQEAMDISKKEMPPTNPIRLGLALNFSVFH
YEIANSPEEAISLAKTTFDEAMADLHTLSEDSYKDSTLIMQLLRDNLTLWTADNAGEEGG
EAPQEPQS
etc...
```

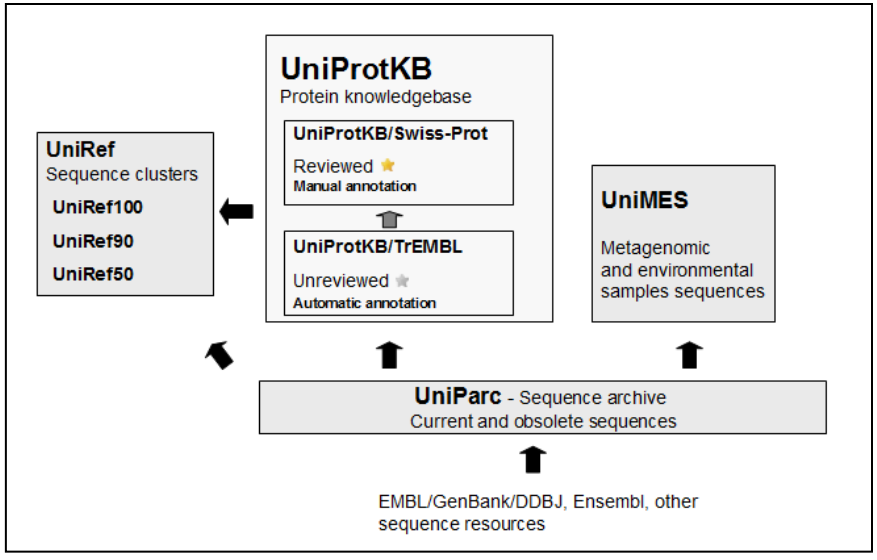
"UniProt (*example*)

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. UniProt is a collaboration between the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). Across the three institutes close to 150 people are involved through different tasks such as database curation, software development and support."

http://www.uniprot.org/help/about

***Example**:*

20,292 entries for taxonomy: "Homo sapiens (Human) [9606]"
AND reviewed:yes



1

PMF Search Parameters Review:

- Trypsin Proteolytic Enzyme and
- 0 Missed Cleave Site (#MC)

Example:
Theoretical Trypsin Digest Peptide Mass List (truncated) for a Single Protein:
Trypsin = Enzyme
Missed Cleave Sites = 0

Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform (PP2A B subunit isoform B'-alpha) (PP2A B subu isoform PR61-alpha) (PR61alpha) (PP2A B subunit isoform R5-alpha)

- Chain Serine/threonine-protein phosphatase 2A at positions 1 - 486 [Theoretical pI: 6.28 / Mw (average mass): 56193.74 / Mw (monoisotopic mass): 56158.15]

mass	position	#MC	modifications	peptide sequence
1806.8538	1-20	0		MSSSSPPAGAASAAISASEK
694.3518	21-26	0		VDGFTR
896.4221	38-45	0	PHOS: 41, 42 1056.3547	SQGSSQFR
1632.8704	46-60	0		SQGSQAELHPLPQLK
1640.7221	61-74	0		DATSNEQQELFCQK
2104.9752	75-92	0		LQQCCILFDFMDSVSDLK
1608.8227	99-112	0		ATLNELVEYVSTNR
1379.7416	113-125	0		GVIVESAYSDIVK
951.5080	126-133	0		MISANIFR
4143.9497	134-168	0		TLPPSDNPDFDPEEDEPTLE ASWPHIQLVYEFFLR
1478.7525	169-181	0		FLESPDFQPSIAK
666.3457	183-187	0		YIDQK
1835.9174	188-202	0		FVQQLLELFDSEDPR
522.2922	205-208	0		DFLK
625.3780	209-213	0		TVLHR
605.3769	218-222	0		FLGLR
506.3085	223-226	0		AFIR
1017.5840	228-235	0		QINNIFLR
3347.7925	236-265	0		FIYETEHFNGVAELLEILGS IINGFALPLK
666.3643	270-274	0		QFLMK
1009.5863	275-283	0		VLIPMHTAK
2150.1466	284-302	0		GLALFHAQLAYCVVQFLEK
1144.6208	303-312	0		DTTLTEPVIR
577.3133	317-320	0		FWPK
566.2603	321-325	0		TCSQK
2479.2676	326-346	0		EVMFLGEIEEILDVIEPTQF K
875.4873	348-354	0		IEEPLFK
1349.6266	359-370	0		CVSSSHFQVAER

http://www.expasy.org/tools/peptide-mass.html

1

PMF Search Parameters:

- Trypsin Proteolytic Enzyme and
- 1 Missed Cleave Site
- MSO = Methionine Oxidation (amino acid modification)

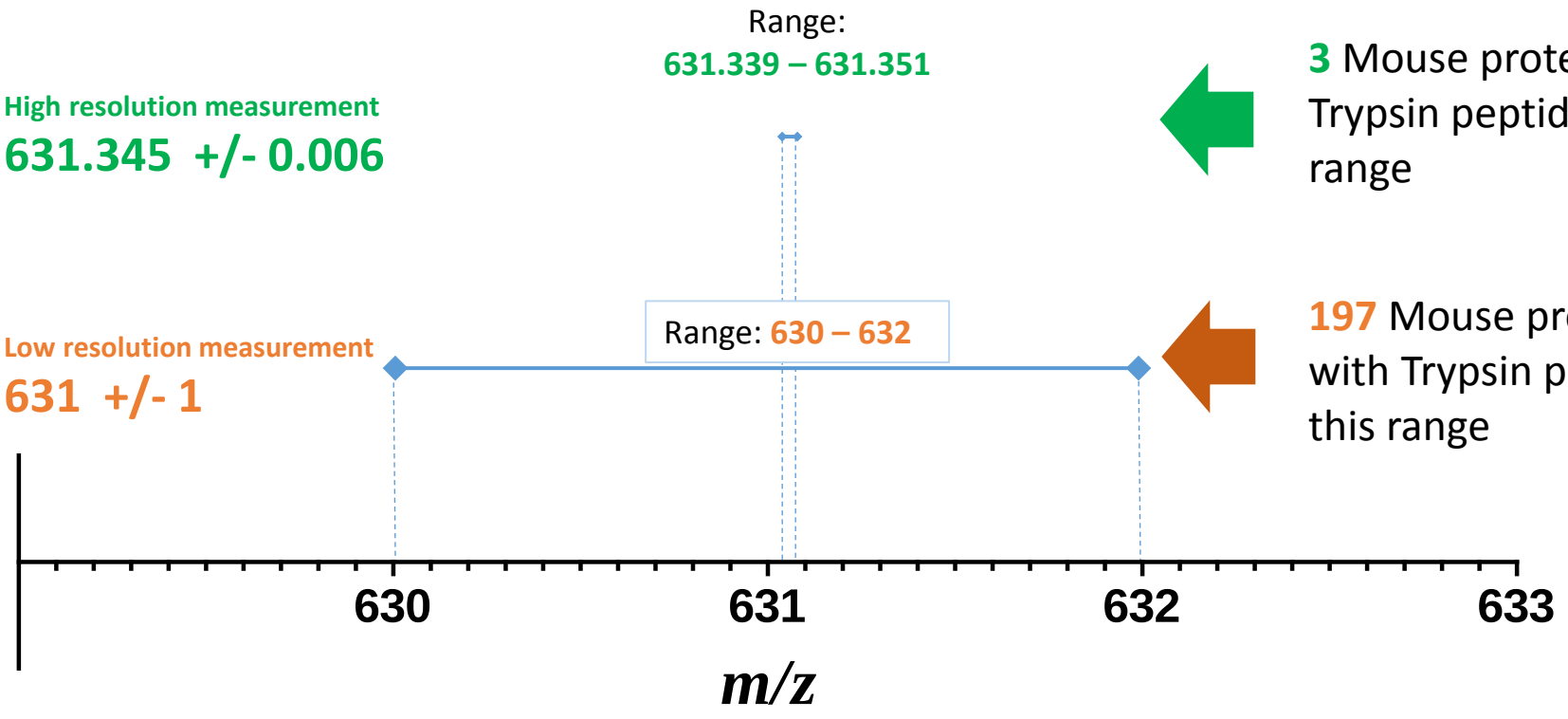
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform (PP2A B subunit isoform B'-alpha) (PP2A B subunit isoform B56-alpha) (PP2A B su isoform PR61-alpha) (PR61alpha) (PP2A B subunit isoform R5-alpha)

- Chain Serine/threonine-protein phosphatase 2A at positions 1 - 486 [Theoretical pI: 6.28 / Mw (average mass): 56193.74 / Mw (monoisotopic mass): 56158.15]

mass	position	#MC	artif modifications		modifications		peptide sequence
1806.8538	1-20	0	MSO: 1	1822.8487			MSSSSPPAGAASAAISASEK
2482.1878	1-26	1	MSO: 1	2498.1827			MSSSSPPAGAASAAISASEK VDGFTR
694.3518	21-26	0					VDGFTR
822.4468	21-27	1					VDGFTRK
502.3096	31-34	1					KAQR
630.3682	32-36	1					AQRQK
1052.5232	37-45	1			PHOS: 41, 42	1212.4558	RSQGSSQFR
896.4221	38-45	0			PHOS: 41, 42	1056.3547	SQGSSQFR
2510.2746	38-60	1			PHOS: 41, 42	2670.2072	SQGSSQFRSQGSQAELHPLP QLK
1632.8704	46-60	0					SQGSQAELHPLPQLK
3254.5746	46-74	1					SQGSQAELHPLPQLKDATSN EQQELFCQK
1640.7221	61-74	0					DATSNEQQELFCQK
3726.6794	61-92	1	MSO: 85	3742.6743			DATSNEQQELFCQKLQQCCI LFDFMDSVSDLK
2104.9752	75-92	0	MSO: 85	2120.9701			LQQCCILFDFMDSVSDLK
2320.1022	75-94	1	MSO: 85	2336.0971			LQQCCILFDFMDSVSDLKSK
604.3664	93-97	1					SKEIK
545.3405	95-98	1					EIKR
1764.9239	98-112	1					RATLNELVEYVSTNR
1608.8227	99-112	0					ATLNELVEYVSTNR
2969.5465	99-125	1					ATLNELVEYVSTNRGVIVES AYSDIVK
1379.7416	113-125	0					GVIVESAYSDIVK
2312.2318	113-133	1	MSO: 126	2328.2267			GVIVESAYSDIVKMISANIF R
951.5080	126-133	0	MSO: 126	967.5029			MISANIFR

http://www.expasy.org/tools/peptide-mass.html

1 PMF Parameter: Peptide Mass Deviation (+/- *m/z*)



1 Mass Measurement Error Calculation

Error Expression	Fractional Error Expression	Multiply by:
ppm	Δ / Theoretical value	1,000,000
%	Δ / Theoretical value	100

where Δ = Experimental (or observed) – Theoretical *m/z* value

Example:

Experimental/observed value (*i.e.*, the data acquired by the mass spectrometer) = 1480.107 *m/z*
Theoretical value (calculated from periodic table, after the peak is identified) = 1480.028 *m/z*
Delta (Δ) = 0.079

Error Expression	Error Equation	Error
ppm	(0.079/1480.028)*1,000,000	53 ppm
%	(0.079/1480.028)*100	0.0053%

A USEFUL REFERENCE:
Brenton AG, Godfrey AR. ***Accurate mass measurement: terminology and treatment of data***. J Am Soc Mass Spectrom. 2010 Nov;21(11):1821-35.

1 MASCOT Peptide Mass Fingerprint Search

http://www.matrixscience.com/

MATRIX SCIENCE

Search this site

Home | Mascot database search | Products | Technical support | Training | News | Blog | Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server

Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

More info

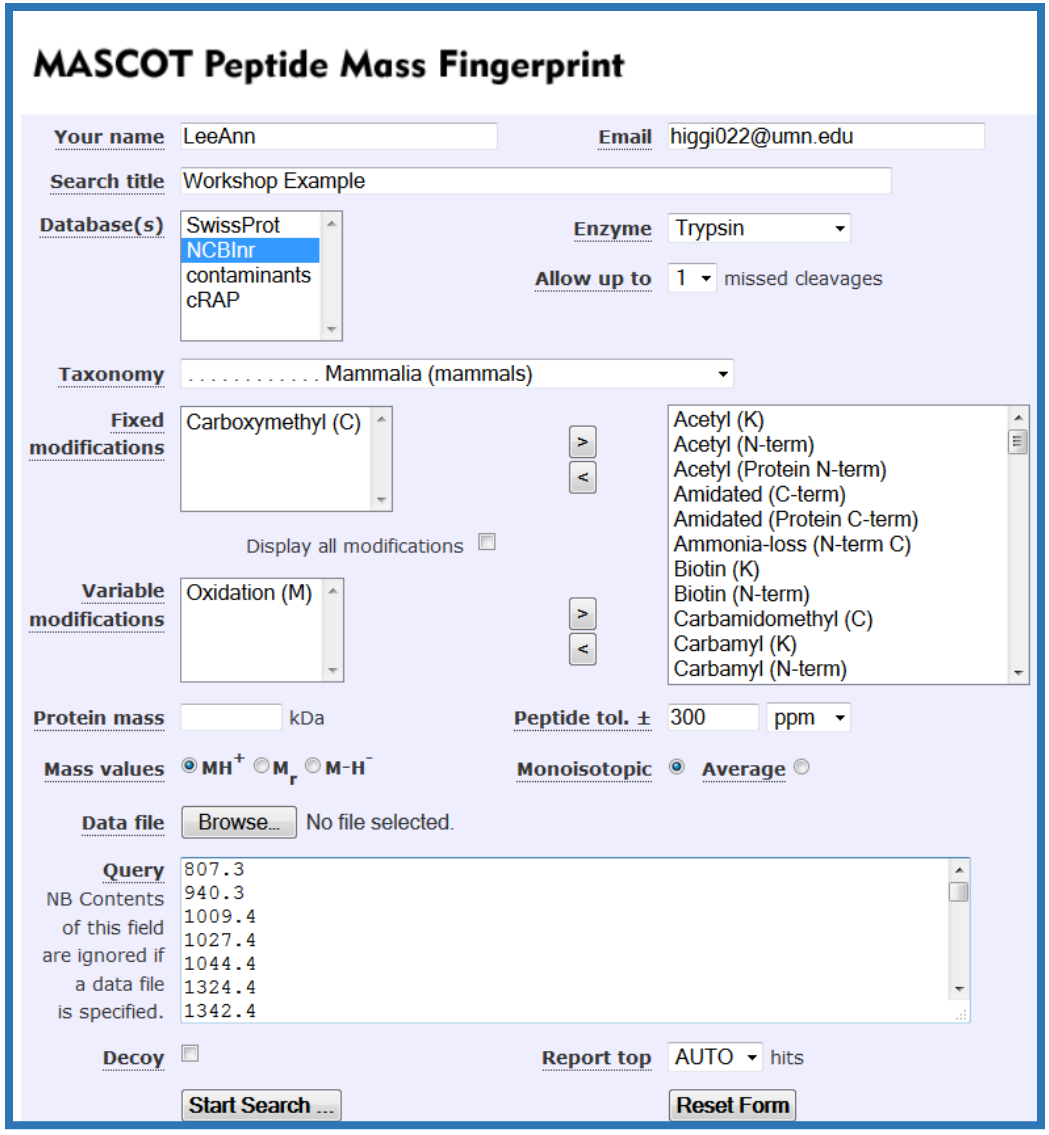
- Mascot overview
- Search parameter reference
- Data file format
- Results report overview

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

Perform search | Example of results report | Tutorial

1 MASCOT PMF Search Parameter Page



MASCOT Peptide Mass Fingerprint

Your name: LeeAnn
Email: higgi022@umn.edu
Search title: Workshop Example
Database(s): SwissProt, NCBInr, contaminants, cRAP
Enzyme: Trypsin
Allow up to: 1 missed cleavages
Taxonomy: Mammalia (mammals)
Fixed modifications: Carboxymethyl (C)
Variable modifications: Oxidation (M)
Display all modifications
Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamidomethyl (C)
Carbamyl (K)
Carbamyl (N-term)
Protein mass: kDa
Peptide tol. ±: 300 ppm
Mass values: MH+ Mr M-H-
Monoisotopic Average
Data file: Browse... No file selected.
Query
NB Contents of this field are ignored if a data file is specified.
807.3
940.3
1009.4
1027.4
1044.4
1324.4
1342.4
Decoy
Report top: AUTO hits
Start Search ...
Reset Form

1 PMF Search

QUERY (Experimental Data)

m/z	Peptide
631.71	NDIAAK
650.71	ELGFQG
662.72	ASEDLK
708.81	TEAEMK
748.90	ALELFR
1272.44	LFTGHPETLEK
1379.69	HGTVVLTALGGILK
1607.81	VEADIAGHGQEVLIR
1817.01	GLSDGEWQQVLNVWGK
1855.06	GHHEAELKPLAQSHATK
3241.65	YLEFISDAIIHVLHSHPGNFGADAQGAMTK

Search

Compare QUERY to THEORETICAL PEPTIDE MASS LIST for each protein in the database

Parameters:
- Enzyme
- Missed cleave site
- Amino acid mods
- Mass tolerance

Score

Probability-based MOWSE Score
(often, the protein with *highest number of peptide matches* has the highest score)

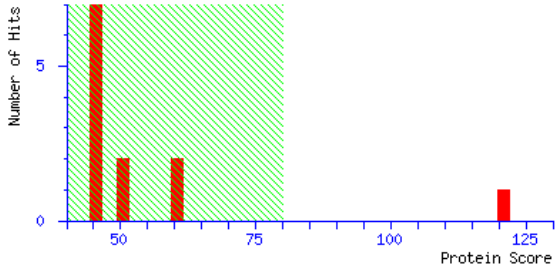
1

MATRIX SCIENCE Mascot Search Results

```
User            : LeeAnn
Email           : higgi022@umn.edu
Search title    :
Database        : NCBInr 20140305 (37425594 sequences; 13257553858 residues)
Taxonomy        : Metazoa (Animals) (5200995 sequences)
Timestamp       : 13 Mar 2014 at 00:03:18 GMT
Top Score       : 121 for gi|30749793, Chain A, K45e Variant Of Horse Heart Myoglobin
```

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 80 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Preferred taxonomy All entries

Re-Search All | Search Unmatched

```
1.    gi|30749793     Mass: 16942     Score: 121     Expect: 4.1e-06  Matches: 9
      Chain A, K45e Variant Of Horse Heart Myoglobin
      gi|7546624      Mass: 16941     Score: 119     Expect: 6.5e-06  Matches: 9
      Chain A, Myoglobin (Horse Heart) Wild-Type Complexed With Co
      gi|255683511    Mass: 17072     Score: 116     Expect: 1.3e-05  Matches: 9
      myoglobin [Equus caballus]
      gi|386783171    Mass: 16967     Score: 116     Expect: 1.3e-05  Matches: 9
      Chain B, Horse Heart Myoglobin: D44kD60K MUTANT WITH ZINC (II) - Deuteroporphyrin Dimethyl Ester
```

1

MATRIX SCIENCE **MASCOT Search Results**

Protein View: gi|30749793

Chain A, K45e Variant Of Horse Heart Myoglobin

Database:	NCBInr
Score:	121
Expect:	4.1e-06
Nominal mass (M_r):	16942
Calculated pI:	6.63
Taxonomy:	Equus caballus

Sequence similarity is available as an NCBI BLAST search of gi|30749793 against nr.

Search parameters

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carboxymethyl (C)
Variable modifications:	Oxidation (M)
Mass values searched:	19
Mass values matched:	9

Protein sequence coverage: 67%

Matched peptides shown in ***bold red***.

```
  1 GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDEFKHLK
 51 TEAEMKASED LKKHGTVVLT ALGGILKKKG HHEAELKPLA QSHATKHKIP
101 IKYLEFISDA IIHVLHSKHP GDFGADAQGA MTKALELFRN DIAAKYKELG
151 FQG
```

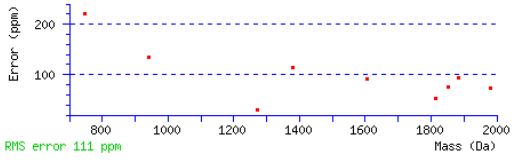
Unformatted sequence string: 153 residues (for pasting into other applications).

Sort peptides by ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Show predicted peptides also

Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
1 – 16	1816.0000	1814.9927	1814.8952	53.8	0	-.GLSDGEWQQVLNVWGK.V
17 – 31	1607.0000	1605.9927	1605.8475	90.5	0	K.VEADIAGHGQEVLIR.L
32 – 42	1271.7000	1270.6927	1270.6557	29.1	0	R.LFTGHPETLEK.F
64 – 77	1379.0000	1377.9927	1377.8344	115	0	K.HGTVVLTALGGILK.K
79 – 96	1982.2000	1981.1927	1981.0493	72.4	1	K.KGHHEAELKPLAQSHATK.H
80 – 96	1854.1000	1853.0927	1852.9544	74.7	0	K.GHHEAELKPLAQSHATK.H
103 – 118	1885.2000	1884.1927	1884.0145	94.6	0	K.YLEFISDAIIHVLHSK.H
134 – 139	748.6000	747.5927	747.4279	221	0	K.ALELFR.N
146 – 153	941.6000	940.5927	940.4654	135	1	K.YKELGFQG.-

No match to: 612.4000, 806.6000, 825.2000, 841.2000, 1329.8000, 1454.8000, 1665.0000, 1723.0000, 1912.2000, 2040.2000

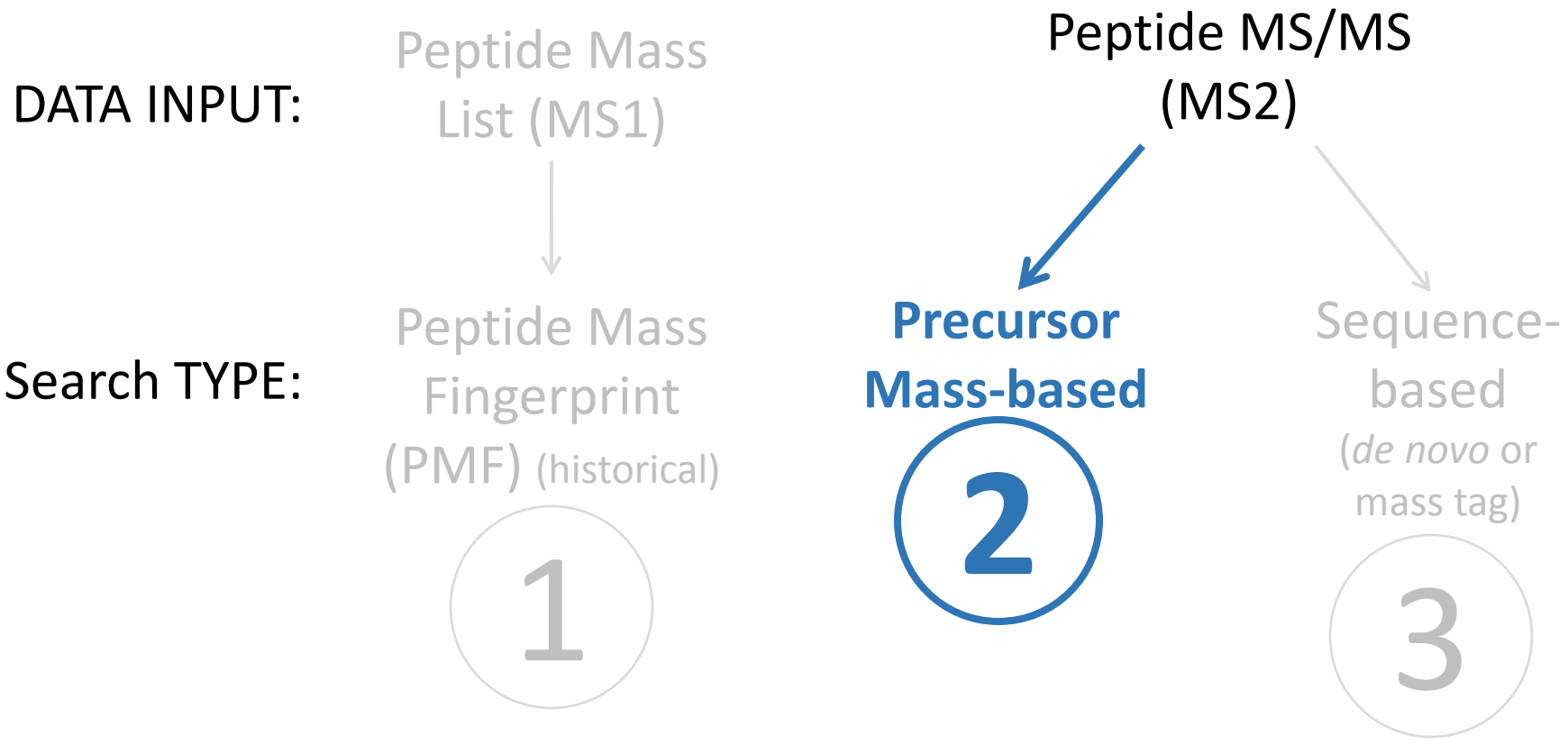


1 PMF Search: Results Interpretation

- Is the protein ID experimentally rational?
- Does the MW of protein in search results match MW determined by SDS-PAGE?
- Does the pI of protein in search results match pI determined by 2D-PAGE?

➢Perform MS/MS if no protein match

Three Strategies for Protein Inference from Peptide MS or MS/MS Data



Tandem MS: Precursor Mass-based Search

Component	Description
Input (Data)	Precursor Mass & Product Ion Masses • Charge state • Average or Monoisotopic
Target	Protein FASTA Sequence Database
Search Parameters	Proteolytic enzyme # Missed enzyme cleave sites Amino acid modifications Mass tolerances: • Peptide/Precursor Mass • Product Ions Masses

2 Tandem MS: Precursor Mass-based Search

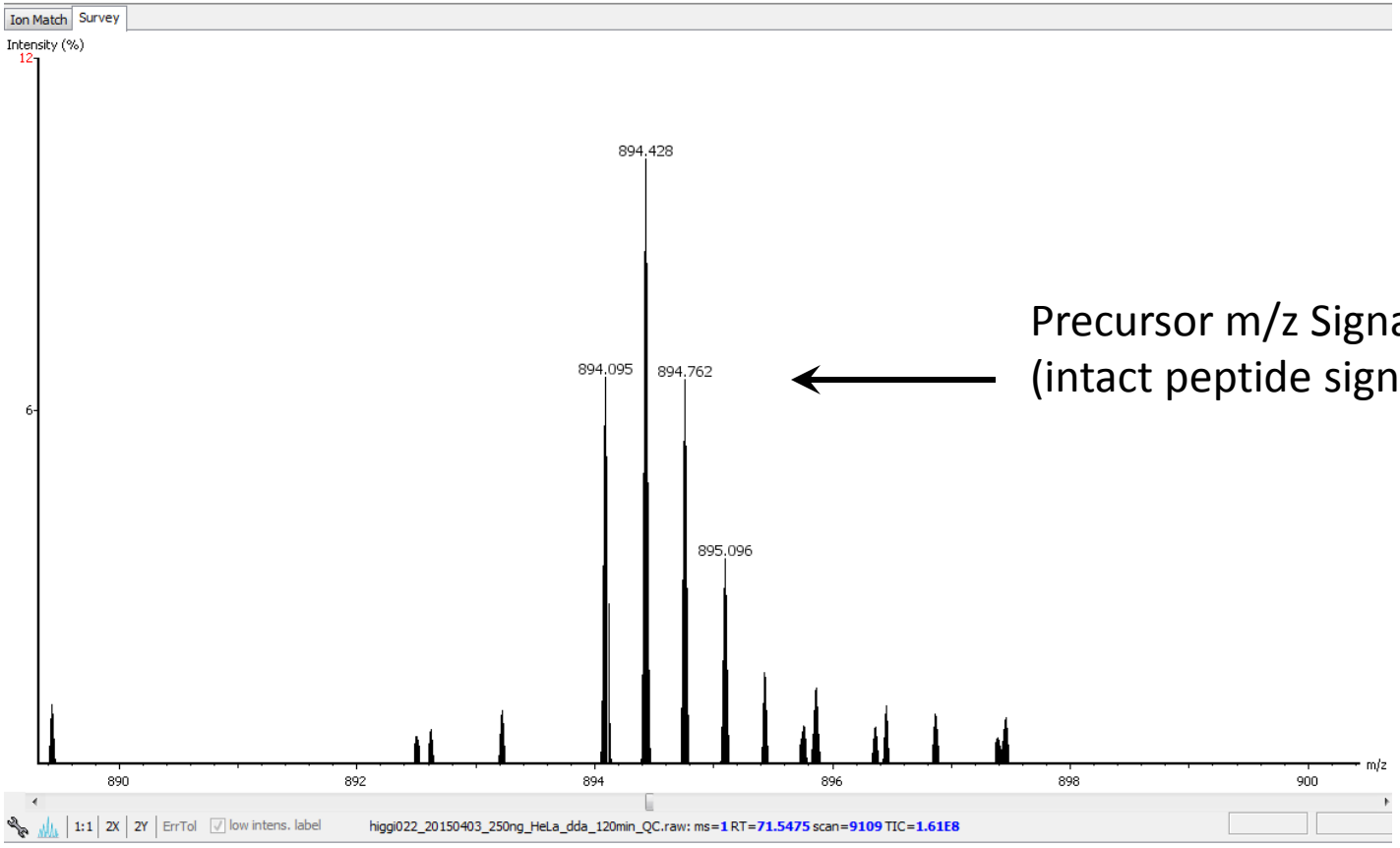
1. Each MS/MS spectra has 2 data-rich Components

2. Database search for each MS/MS spectra has 2 STEPS

2 Tandem MS: Precursor Mass-based Search

Calculate Precursor Mass from Precursor *m/z* Signal

- Precursor *m/z* (monoisotopic) = 894.0941
- Precursor Charge State = 3
- Precursor Mass (Da) = 2679.2559



Precursor m/z Signal
(intact peptide signal)

2 Database search for Peptides

- Target peptide mass = precursor mass
  - example = 2679.2559
- Target peptide mass range = precursor mass + / - precursor mass tolerance (user-specified)
  - example = 50 ppm mass tolerance
  - Example = 2679.2559 +/- 0.1340
  - Example = 2679.1219 – 2679.3899

Peptide Candidate Matches for 2679.2559 precursor mass

Mr(expt)	Mr(calc)	ppm	Peptide Candidate
2679.2559	2679.3798	-46.2	GAVQVFIMLLLATVSDCAVITGACER
2679.2559	2679.1846	26.6	SYDPPCPGHWTPEAPGSGTTCPGLPR
2679.2559	2679.2889	-12.3	DGETAEEQGGPVPPPVAPGGPGLGGAPGGR
2679.2559	2679.3286	-27.1	GAGALALGASEPCNLSSAAPLPDGAATAAR
2679.2559	2679.2010	20.5	AVDFQEAQSYADDNSLLFMETSAK
2679.2559	2679.3102	-20.3	GTYLPQTYIIQEEMVVTEHVSDK
2679.2559	2679.25575	0.1	ERPTPSLNNNC(carbamidomethyl)TTSEDSLVLYNR

… and many more…

2 Tandem MS: Precursor Mass-based Search

Compare theoretical product ions to experimental product Ions from MS2 spectrum

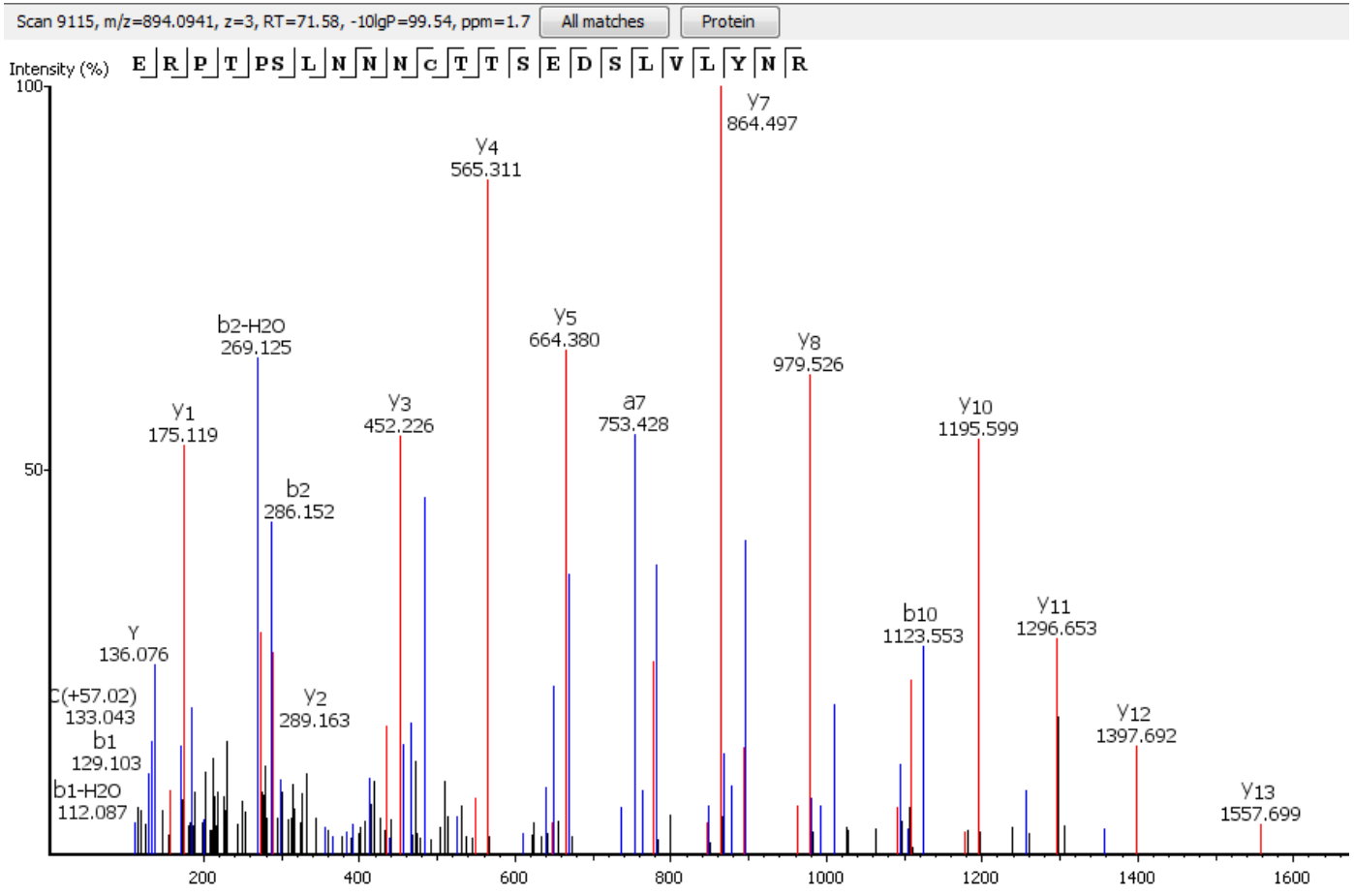
2 COMPARE

theoretical product ions (table below) to experimental product Ions from MS2 spectrum (matched ions are colored blue and red)

Ion Match | Survey

#	Immonium	b	b-H2O	a	a-H2O	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	102.055	130.087	112.087	102.055	84.045	65.525	E					23
2	129.103	286.152	268.142	258.156	240.146	143.576	R	2551.220	2533.210	2534.193	1276.110	22
3	70.066	383.189	365.189	355.214	337.199	192.102	P	2395.119	2377.109	2378.092	1198.060	21
4	74.061	484.253	466.241	456.258	438.243	242.626	T	2298.067	2280.056	2281.040	1149.533	20
5	70.066	581.305	563.294	553.310	535.299	291.152	P	2197.019	2179.009	2179.992	1099.010	19
6	60.045	668.338	650.329	640.354	622.331	334.668	S	2099.966	2081.956	2082.939	1050.483	18
7	86.097	781.423	763.413	753.428	735.428	391.194	L	2012.934	1994.924	1995.907	1006.967	17
8	87.056	895.468	877.454	867.478	849.459	448.232	N	1899.850	1881.840	1882.823	950.425	16
9	87.056	1009.513	991.520	981.519	963.501	505.253	N	1785.807	1767.797	1768.780	893.379	15
10	87.056	1123.553	1105.556	1095.559	1077.544	562.275	N	1671.764	1653.754	1654.737	836.382	14
11	133.043	1283.580	1265.570	1255.599	1237.575	642.290	C(+57.02)	1557.699	1539.711	1540.694	779.361	13
12	74.061	1384.628	1366.617	1356.616	1338.622	692.814	T	1397.692	1379.680	1380.664	699.345	12
13	74.061	1485.676	1467.665	1457.681	1439.670	743.338	T	1296.653	1278.632	1279.616	648.821	11
14	60.045	1572.708	1554.697	1544.713	1526.702	786.854	S	1195.599	1177.585	1178.581	598.298	10
15	102.055	1701.750	1683.740	1673.755	1655.745	851.375	E	1108.568	1090.577	1091.536	554.782	9
16	88.040	1816.777	1798.767	1788.782	1770.772	908.889	D	979.526	961.510	962.491	490.260	8
17	60.045	1903.809	1885.799	1875.814	1857.804	952.405	S	864.497	846.483	847.478	432.747	7
18	86.097	2016.893	1998.883	1988.898	1970.888	1008.947	L	777.464	759.451	760.435	389.231	6
19	72.081	2115.962	2097.951	2087.967	2069.956	1058.481	V	664.380	646.367	647.351	332.689	5
20	86.097	2229.046	2211.035	2201.051	2183.040	1115.023	L	565.311	547.299	548.286	283.155	4
21	136.076	2392.109	2374.099	2364.114	2346.104	1196.555	Y	452.226	434.215	435.199	226.613	3
22	87.056	2506.152	2488.141	2478.157	2460.146	1253.576	N	289.163	271.151	272.137	145.081	2
23	129.114						R	175.119	157.099	158.092	88.059	1

2 Evaluate Spectrum for Peptide Candidate Match Quality



2 Tandem MS: Precursor Mass-based Search

SUMMARY- for Each MS/MS Spectrum:

- Generate **theoretical** product ion tables for all peptide candidates
- Compare **theoretical** product ions to **experimental** product Ions from MS2 spectrum
- **Score** all candidate peptides
- **Rank** peptides by Score

2 Tandem MS: Precursor Mass-based Search

STEP 1

Find **Theoretical** Peptide Matches from *in silico* Protein Digest in the range: "**Experimental** Precursor Mass +/- Mass Tolerance"

STEP 2

For each candidate peptide from Step 1: Compare **Theoretical** Product Ions (b & y, etc) to **Experimental** Product ions (data)

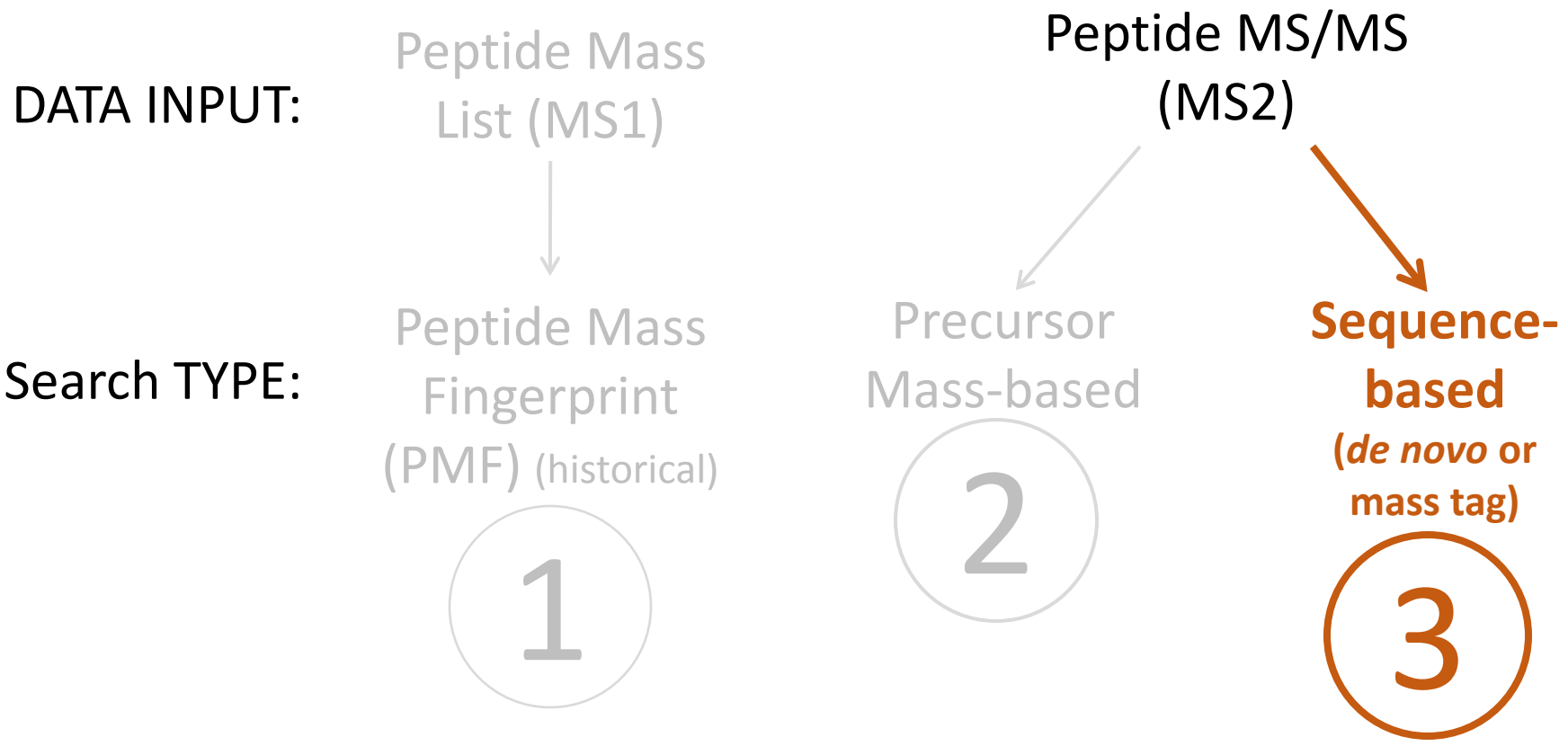
STEP 3

SCORE: many software programs/ algorithms
Peptide rank
PROTEIN Report and Grouping: many variations

2 Tandem MS: Precursor Mass-based search Software at UM

- Sequest
- X!Tandem
- MASCOT

Three Strategies for Protein Inference from Peptide MS or MS/MS Data



3

MS/MS Sequence-based Software at UM

- ProteinPilot® (AB Sciex)
- PEAKS (BSI)

Open-source Software

- Direct Tag (D Tabb Lab)

3 Sequence-based Searches: WHEN?

- Unsequenced genome (protein(s) of interest are not in the database) → HOMOLOGY-based search
- Search for amino acid MUTATIONS
- Search for LARGE NUMBERS of Post Translational Modifications (PTM's)

3 Tandem MS: Sequence-based Search

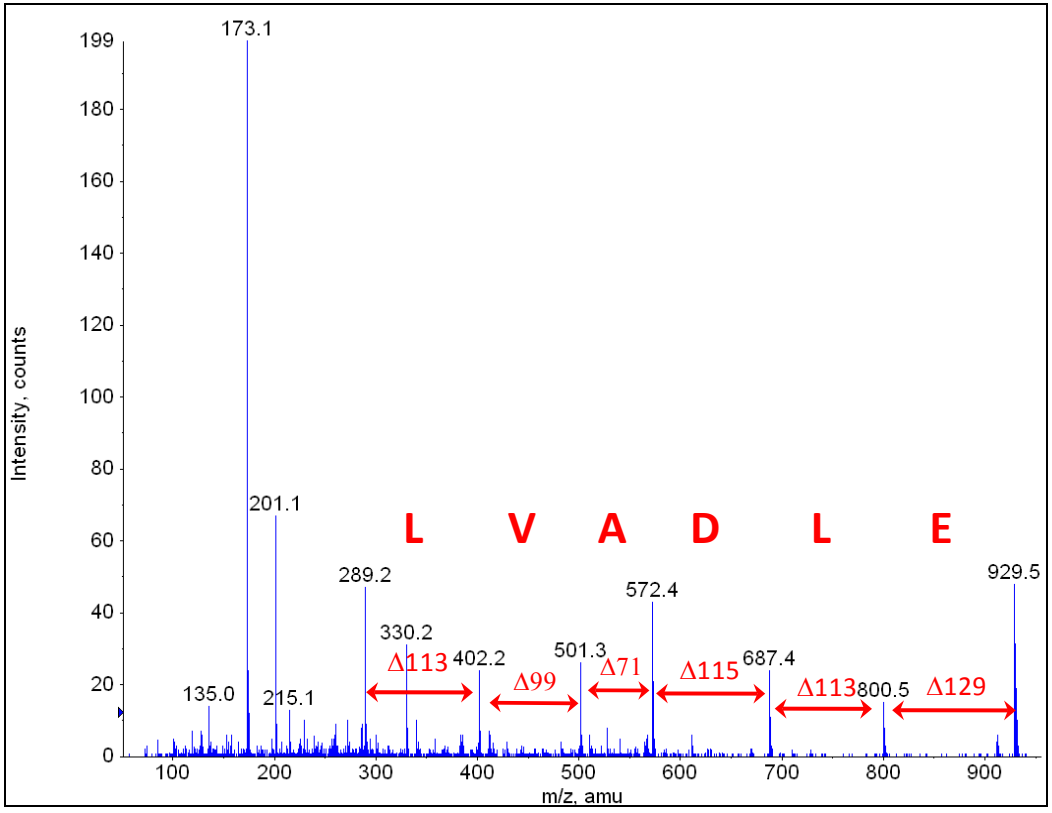
STEP 1:

Obtain [full or partial] amino acid sequence string directly from the spectrum by ***de novo*** Sequencing

de novo (Latin) "from the beginning"

3 Tandem MS: Sequence-based Search

STEP 1: *de novo* sequencing: amino acid sequence is determined from delta mass values for a series of successive peptide b- or y-type product ions from a Product Ion (MS/MS) spectrum



EXAMPLE *de novo* sequencing:

"delta masses" between peaks correspond to amino acid residue masses

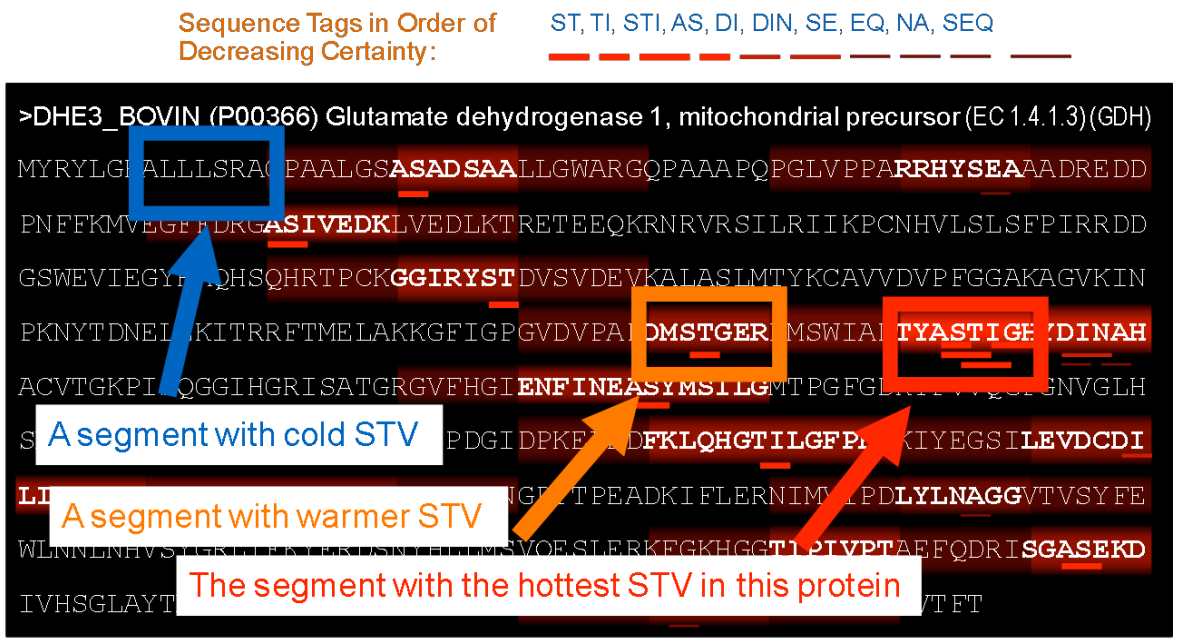
See amino acid residue masses: http://www.ionsource.com/Card/aatable/aatable.htm

3 Tandem MS: Sequence-based Search

STEP 2: QUERY THE PROTEINS

Example: AB Sciex Paragon® Algorithm

For each MS/MS spectrum, multiple, short sequences (taglets) are generated, scored, and matched against protein sequences. "Sequence Temperature Values" (STV) rank candidate amino acid sequence segments. STV is related to the number of taglet hits per region.



Slide content: courtesy of Sean L. Seymour, AB Sciex

3 Tandem MS: Sequence-based Search

STEP 3: Candidate peptide sequences within protein sequences are identified with Precursor mass AND sequence tag information

Include:

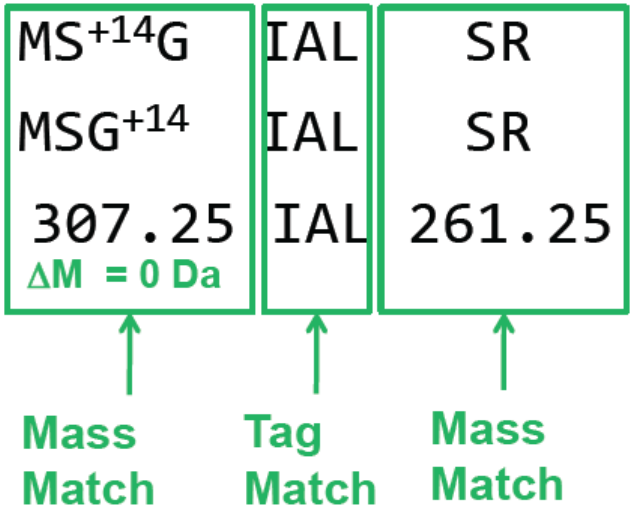
- Enzyme specificity
- Mass tolerance
- Amino acid modifications
- Amino Acid mutations

STEP 4: Candidates are Ranked and SCORED

3 Tandem MS: Sequence-based Search

STEP 3b: Error-tolerant Sequence Tag Reconciliation Includes Amino Acid Mutation Search Component

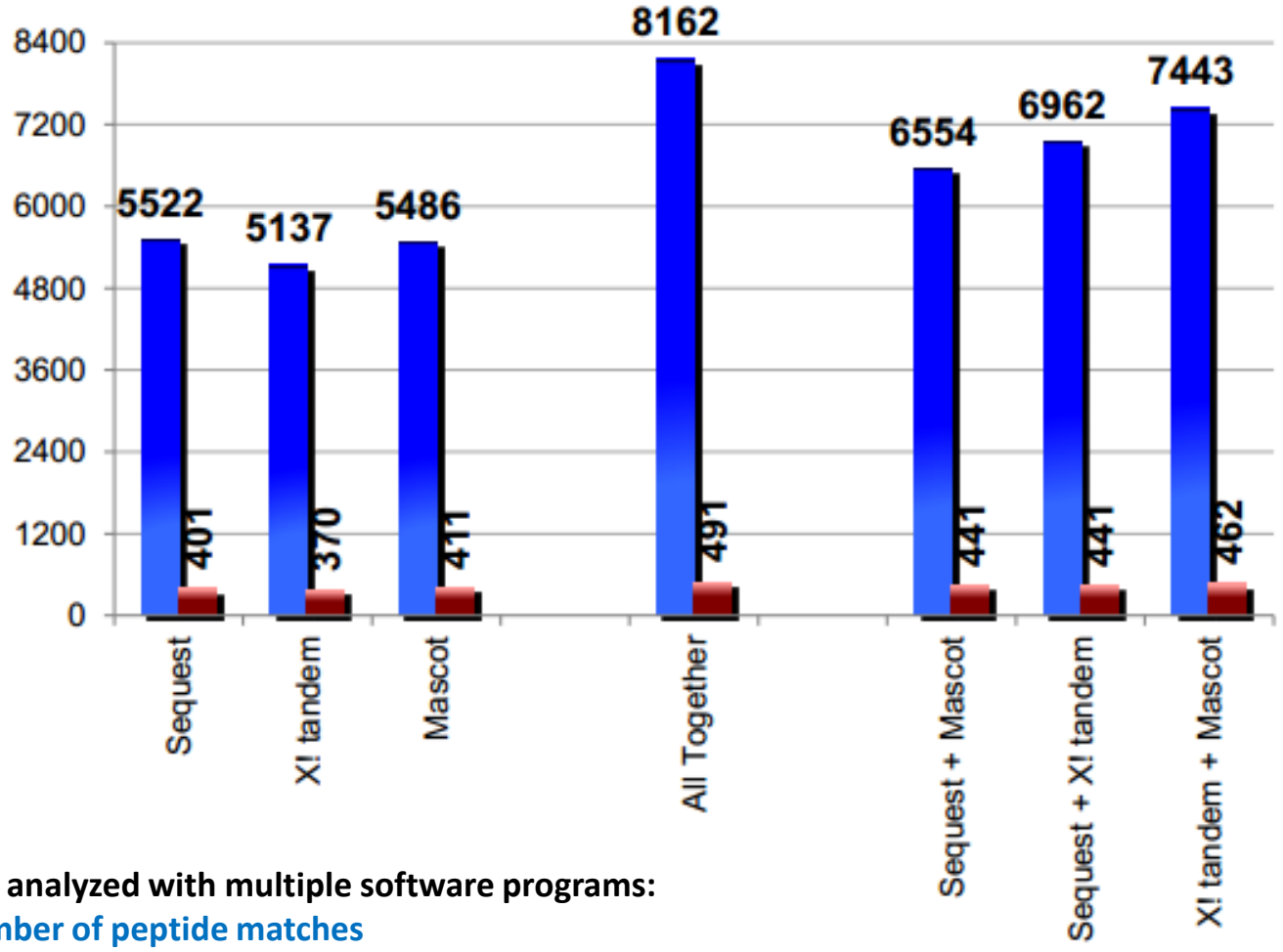
- Only some residues can be mutated to explain mass shift.
- Ser→Thr: $MS^{+14}G$ IAL SR
- Gly→Ala: MSG^{+14} IAL SR
- Spectrum: 307.25 IAL 261.25



Combination: sequence tag & mass match

Slide content: adapted from David L. Tabb, Vanderbilt University: Bioinformatics 2009, Baltimore MD

Maximize Identifications with Multiple Software Programs



Single Dataset analyzed with multiple software programs:

- **Blue = number of peptide matches**
- **Red = number of protein matches**

Search Algorithm Conclusions

1) Several different search algorithm types exist
2) Each program provides useful results - no unified method exists
3) The same data searched using different algorithms may yield different results (scoring and ranking schemes are distinct)