Mass Spectrometric Analysis of Proteins and Peptides

Outline

1. The fundamentals—brief overview of hardwares and data acquisition.
2. Protein characterization, i.e., sequencing.
3. Protein identification.
4. Protein quantification.
5. Other applications.

Tell me, my dear, what is proteome and it's origin??

Proteome

The PROTEin complement of a genOME

"Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How to Do It"

Biotech. Gen. Eng. Rev. (1995) 13, 19-50.
Wilkins, MR; Sanchez, JC; Gooley, AA; Appel, RD; Humphrey-Smith, I; Hochstrazzer, DF and Williams, KL.

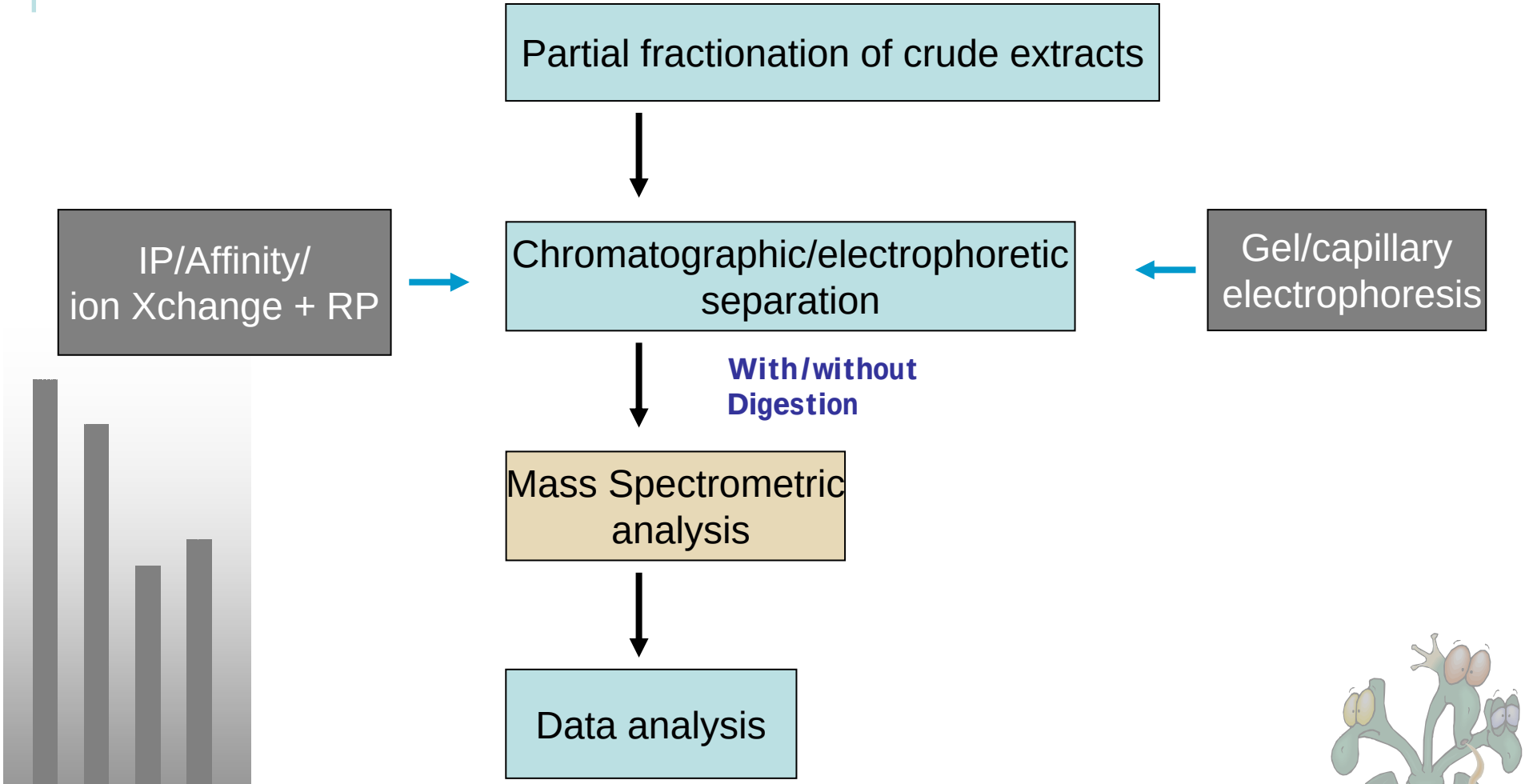
- "Tryptic Mapping of Recombinant Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry"

  Billeci & Stults (1993) *Anal. Chem.* 65, 1709

- "Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases"

  Henzel et al. (1993) *Proc. Natl. Acad. Sci. USA* 90, 5011

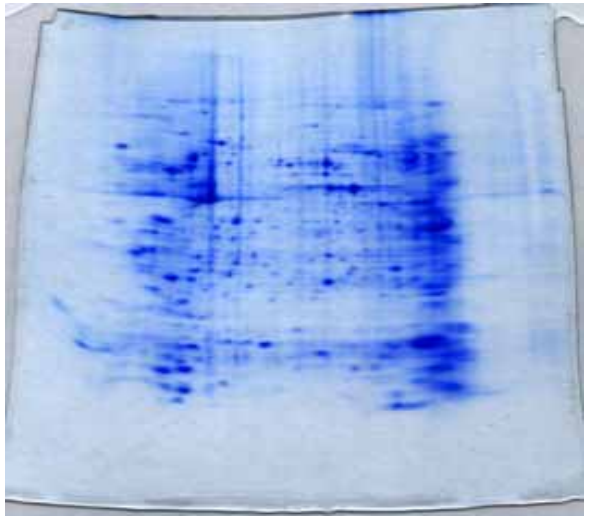
General procedure



Gel separation

IEF/SDS PAGE

pH 3 - 10 NL, 18 cm strip



Angelika Gorg's manuals
http://www.weihenstephan.de/blm/deg/

Do I need to go through IEF?

Read O'Farrell, P.H. (1975)
J. Biol. Chem. 250, 4007-4021.

IEF in tube gels vs IPG
Equilibrium vs non-equilibrium IEF

Usual Problems:

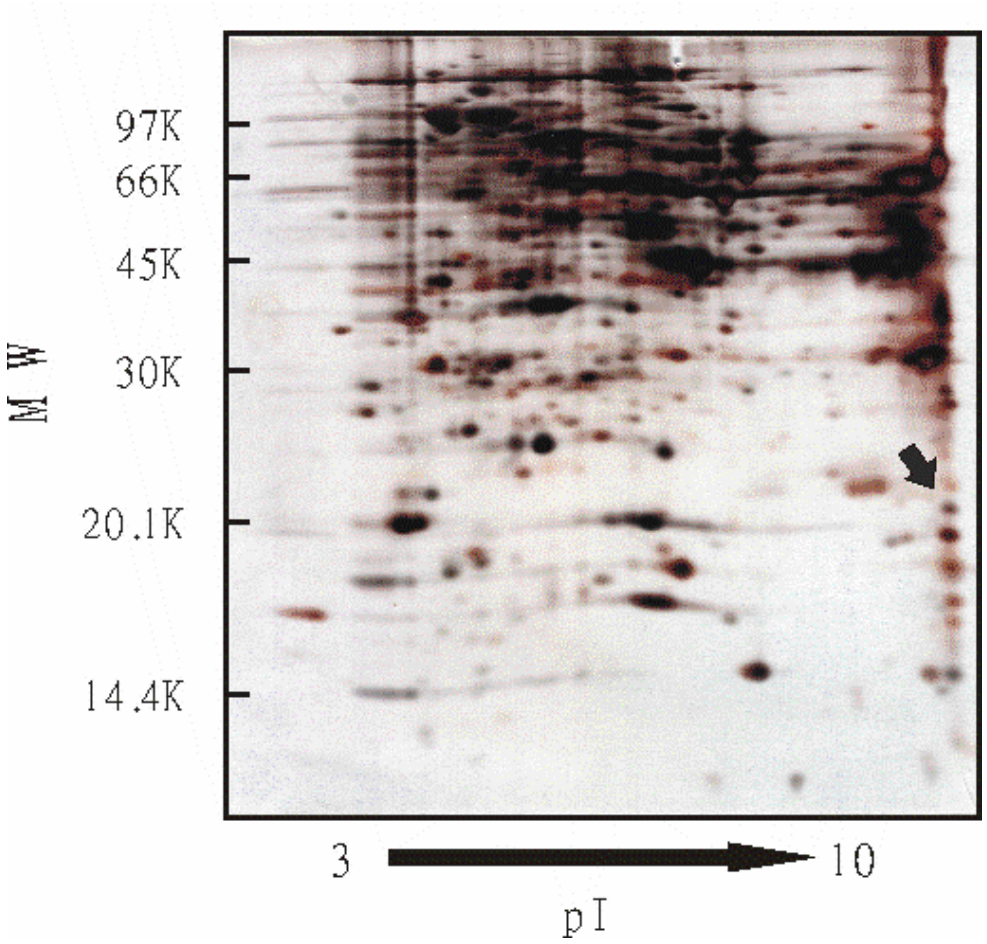
Sample load volume vs concentration

3 mm tube, 1 cm = 70 ul

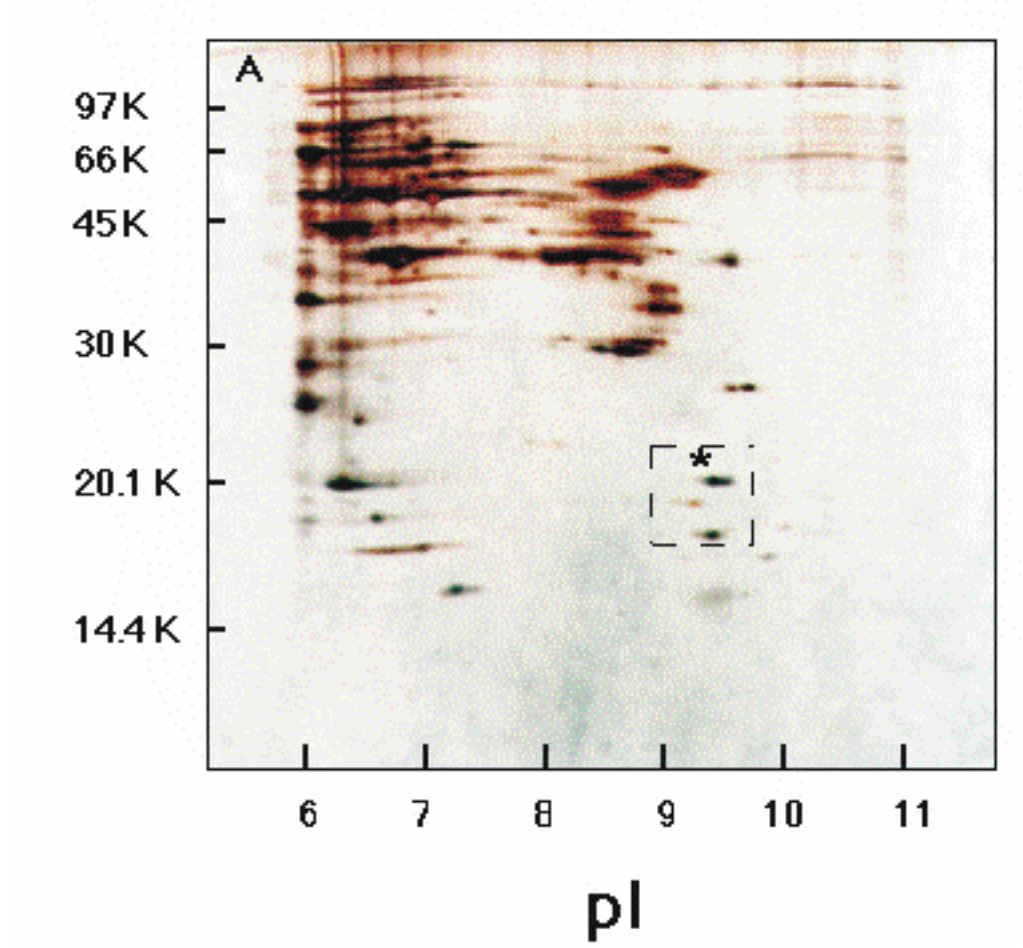
With 1 mg protein loading, conc = 14.3 mg/ml

Solubility, cytosolic vs membrane proteins, interfering materials salt, nucleic acids, lipids

Avoid pH extremes



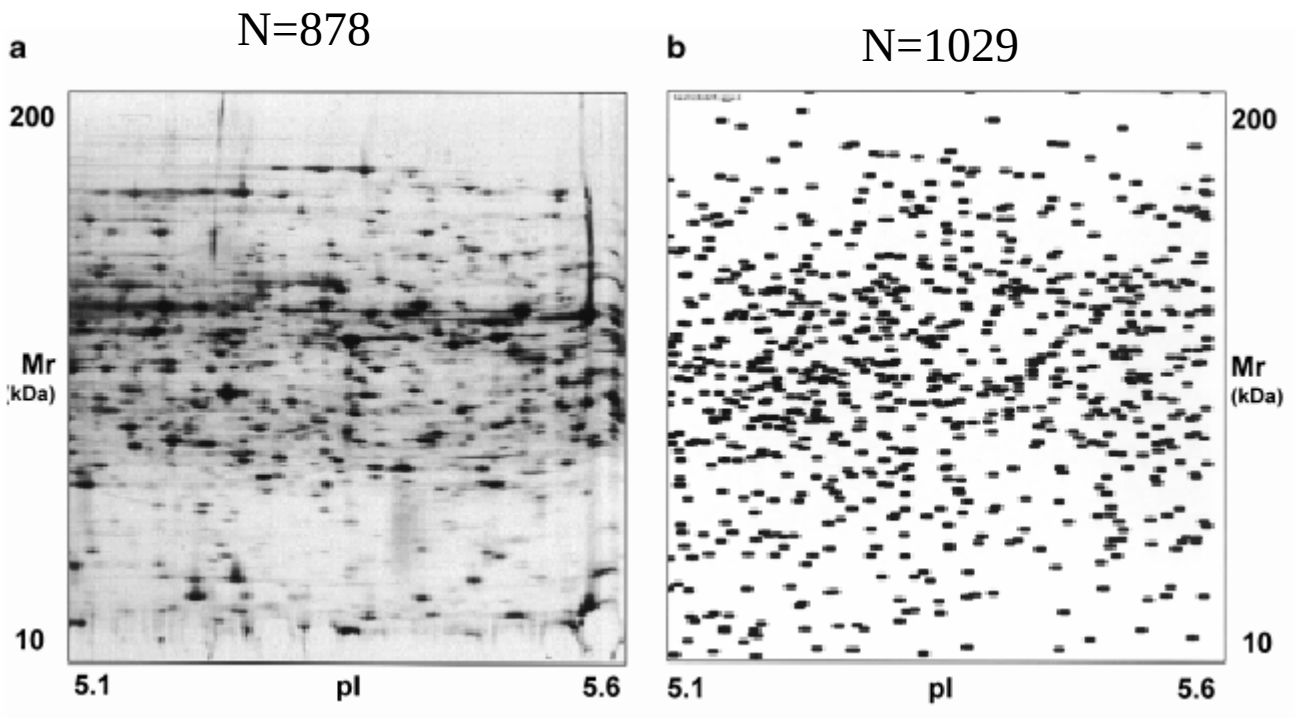
Same sample, different range!
You see part of the total at all time!



2D gel cannot see all the proteins!

- With 3-10, 18 cm strip, 2.57 cm per pI unit.
- With 6-11, 18 cm strip, 3 cm per pI unit.
- Short Range (around neutral pH), 24 cm gel strips.
- 18 cm 2nd-D, assume 500 daltons separated by 2 mm, from top to bottom, covers only 45,000 daltons.
- Gradient vs linear polyacrylamide gel.

- 2D gel cannot see all the proteins!



E. Coli proteins

Corthals et al. Electrophoresis (2000) 21, 1104-1115

- Acrylamide-agarose copolymers: improved resolution of high molecular mass proteins in two-dimensional gel electrophoresis

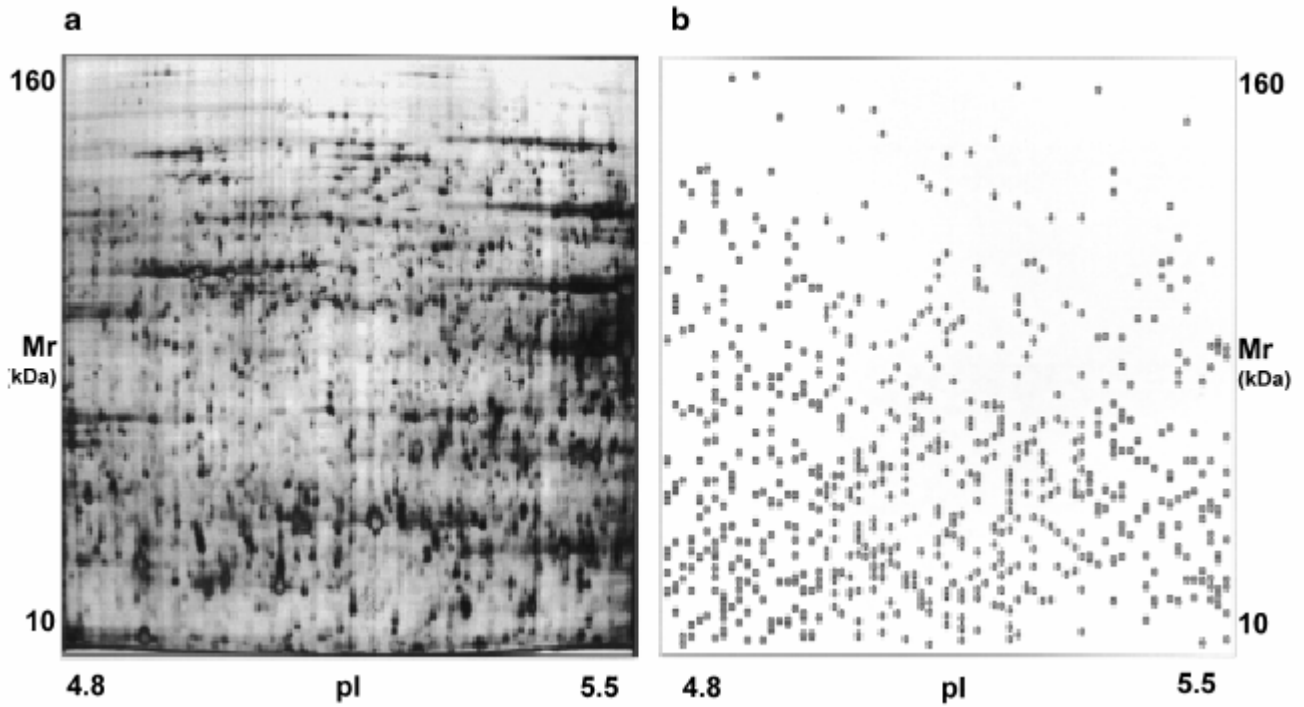
  Roncada et al. (2005) Proteomics 5, 2331-2339.

- An agarose-acrylamide composite native gel system suitable for separating ultra-large protein complexes

  Suh et al. (2005) Anal. Biochem. 343, 166-175.

2D gel see too many proteins!

Degraded/modified samples (multiple spots/protein)

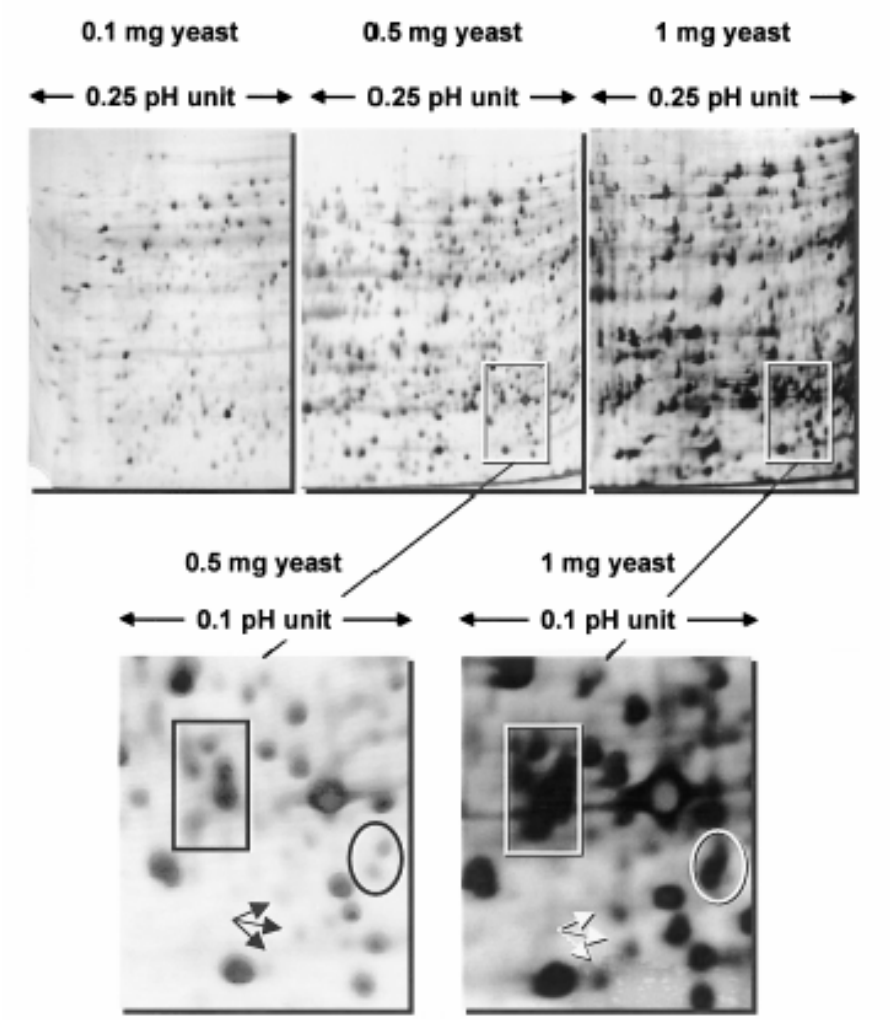


N= 2700, 500 ug

N=852

Yeast proteins

Boss! How much protein should I load?



Gygi et al. *Proc. Natl. Acad. Sci. USA* 2000, 97: 9390-9395.

"Evaluation of two-dimensional gel electrophoresis based proteome analysis technology"

Standard procedure

Gel separation

Isolate spots

You have to see the spots first!

Visualization:

Coomassie stain,

Colloidal blue stain,

Neuhoff et.al., Electrophoresis, 1985, 6, 427-448.

Blue-silver stain,

Electrophoresis (2004) 25, 1327-1333.

Silver stain,

www.narrador.emble-heidelberg.de/GroupPages/PageLink/activities/protocols.html

Sypro dyes

I f U have the money!

Negative stain for not too complicated patterns.

C. Fernandez-Patron et al. (1995) *Anal. Biochem*. 224, 203-211.

Gillespie & Elliott (2005) Anal. Biochem. 345, 158-160.

“Comparative advantage of imidazole—sodium dodecyl sulfate-Zinc reverse staining in polyacrylamide gel”

Standard procedure

Gel separation

Isolate spots

Enzyme digestion

Many steps are no longer needed!

Optimized Sample-Processing Time and Peptide Recovery For the Mass Spectrometric Analysis of Protein Digests

Terry, DE, Umstot, E. and Desiderio, DM
J Am Soc Mass Spectrom 2004, 15, 784-794

Usual considerations:

- Most of the time use modified trypsin.
- NH_4HCO_3 is the usual buffer. Why?
- Don't need too much trypsin, 0.5 uM.
- Need Ca^{++} for the reaction, ~500-fold of the enzyme or 250 uM.
- Excess Ca^{++}, what would happen?
- Temperature? Depends on whether you want internal standards.
- Additives in the buffer?

"Fast-response proteomics by accelerated in-gel digestion of proteins"

Havlis et al. (2003) Anal. Chem. 75, 1300-1306.

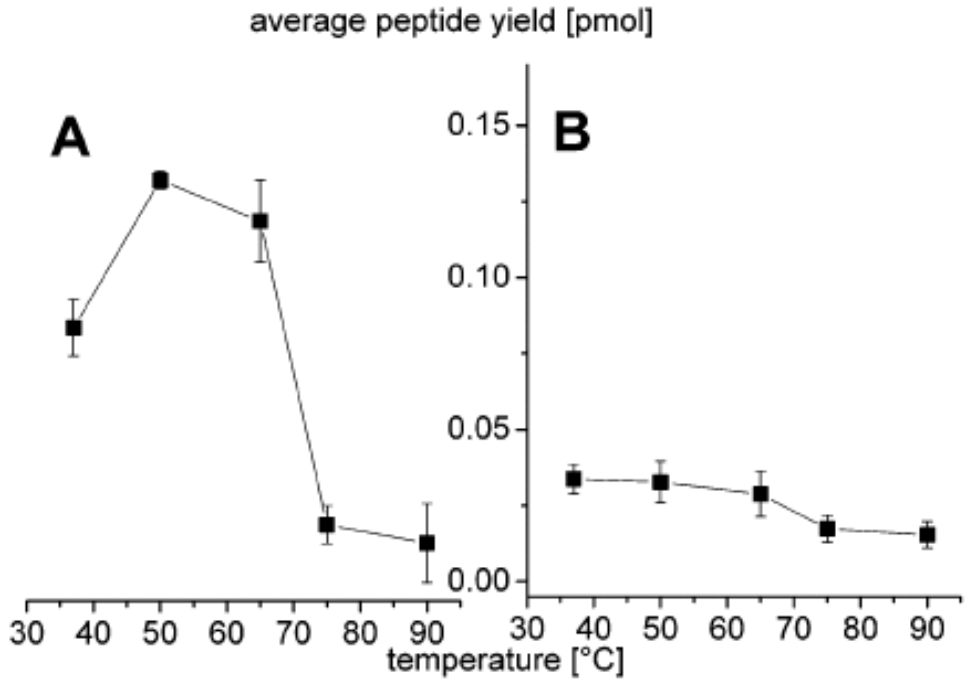


Figure 1. Effect of digestion temperature on the peptide yield: (A) modified trypsin; (B) native (unmodified) trypsin. The enzyme concentration was 0.5 μM and digestion time 30 min. Average peptide yield of control digestion (CDP) was 0.25 pmol.

Kinetic characterization of sequencing grade modified trypsin

Finehout et al. Proteomics (2005) 2319-2321.

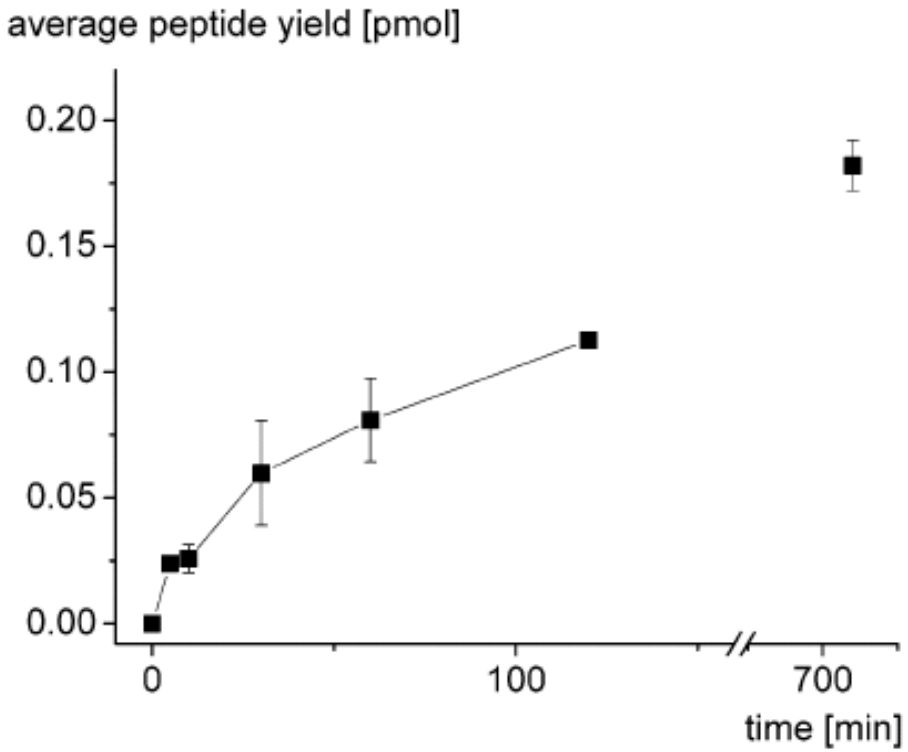


Figure 2. Time curve of digestion by the modified trypsin. The enzyme concentration was 0.5 μM and digestion temperature 58 °C. The peptide yield in the control experiment (CDP) was 0.17 pmol.

You smart people, can you see what is wrong with the data when comparing this graph with the last one???

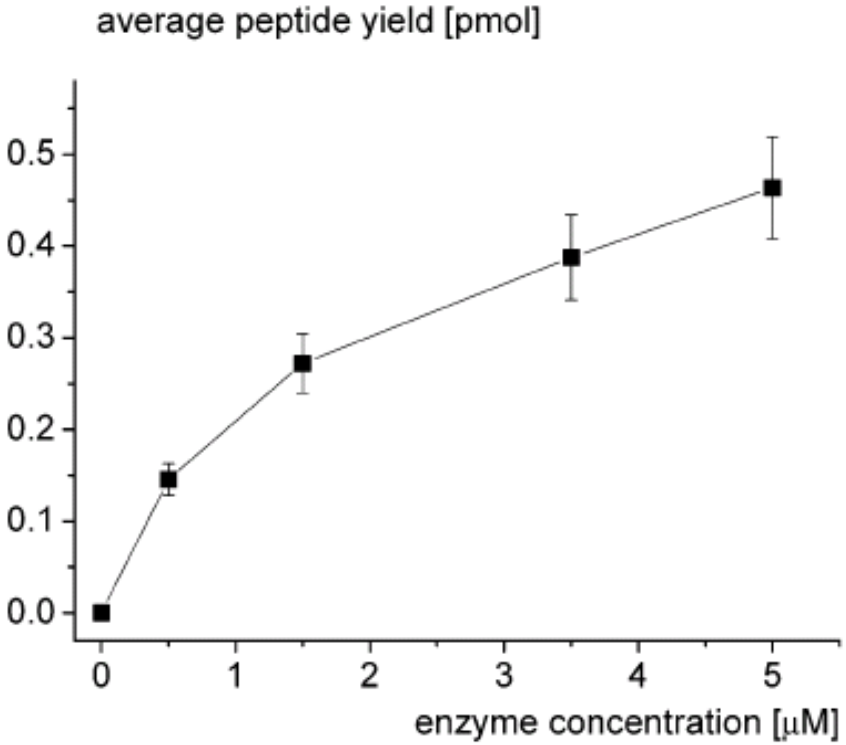


Figure 3. Effect of trypsin concentration on the peptide yield. Digestion was performed with the modified trypsin at 58 °C for 30 min. Average peptide yield of the control digestion (CDP) was 0.21 pmol.

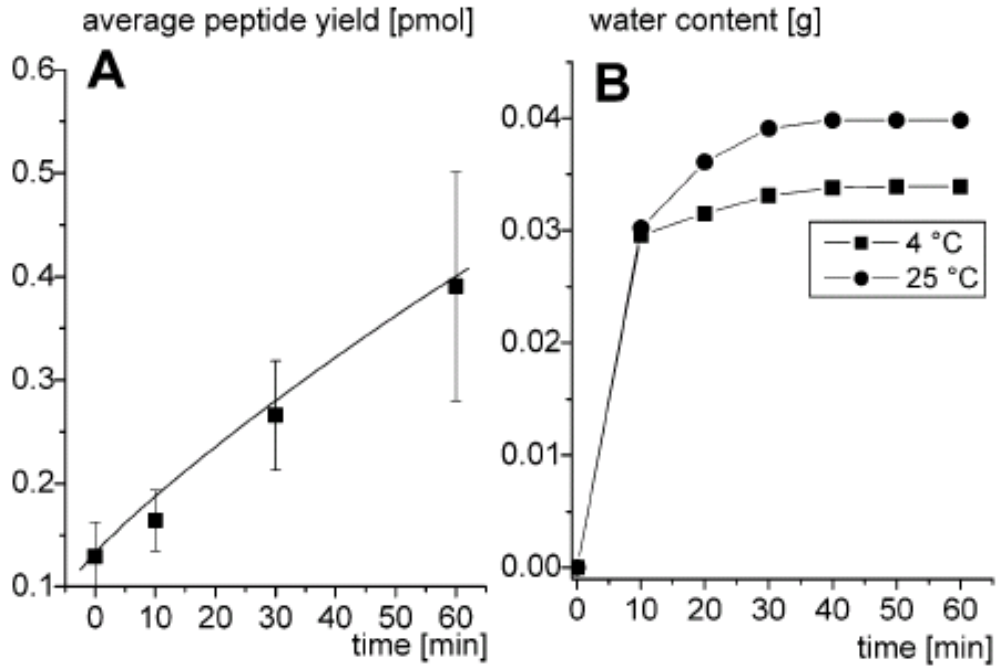


Figure 4. Saturation of dehydrated gel pieces with the trypsin solution and its effect on the peptide yield. (A) Peptide yield versus on rehydration time. (B) Amount of digestion buffer absorbed by gel pieces at 4 and 25 °C. Total weight of dry gel pieces was 5.2 mg.

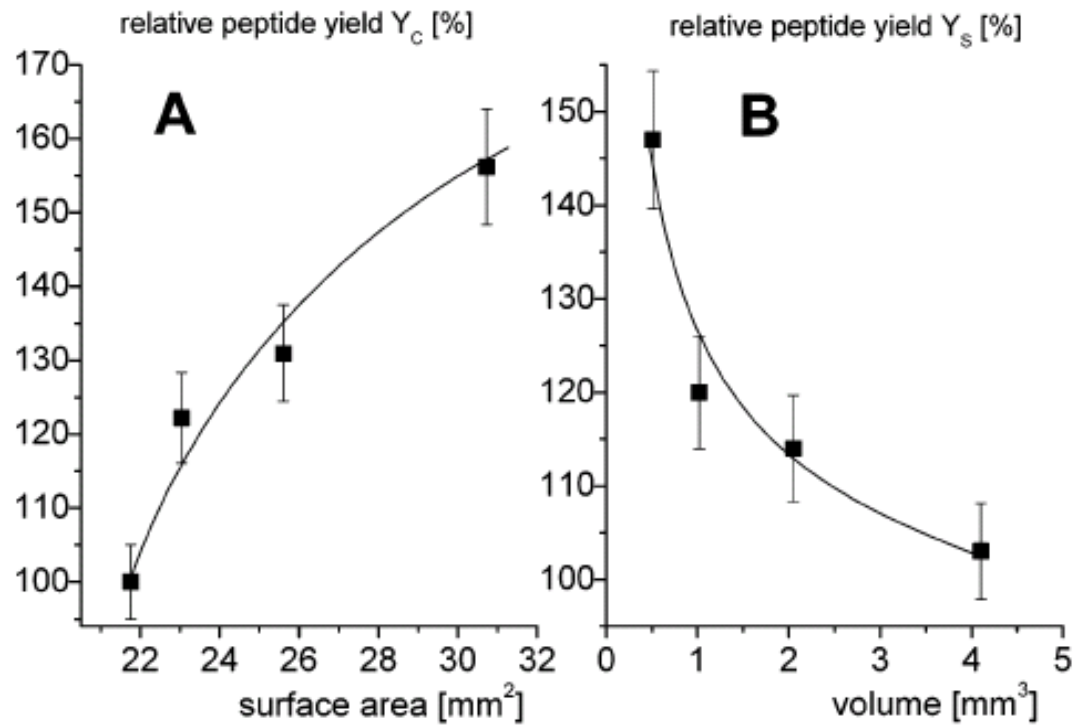


Figure 5. Effect of gel cutting (A) and gel dehydration (B) on the digestion yield. (A) Relative peptide yield versus the surface area of gel pieces, $Y_S = (Y_X/Y_1) \times 100\%$, where Y_X is the yield from the band cut in X equal pieces; Y_1 is the yield from the uncut band (reference). (B) Relative peptide yield from dehydrated gel pieces, compared to the yield from nondehydrated pieces of the same size, $Y_C = (Y_D/Y_{ND}) \times 100\%$, where Y_D is the yield from the gel pieces dehydrated prior to soaking in trypsin solution and Y_{ND} is the yield from nondehydrated gel pieces of the same size (reference). Total internal volume of gel pieces was ~4 mm^3.

Peptide extraction

- Can the peptides diffuse out?
- Solvents used?

- Microwave digestion & extraction

  Zhong et al. (2005)

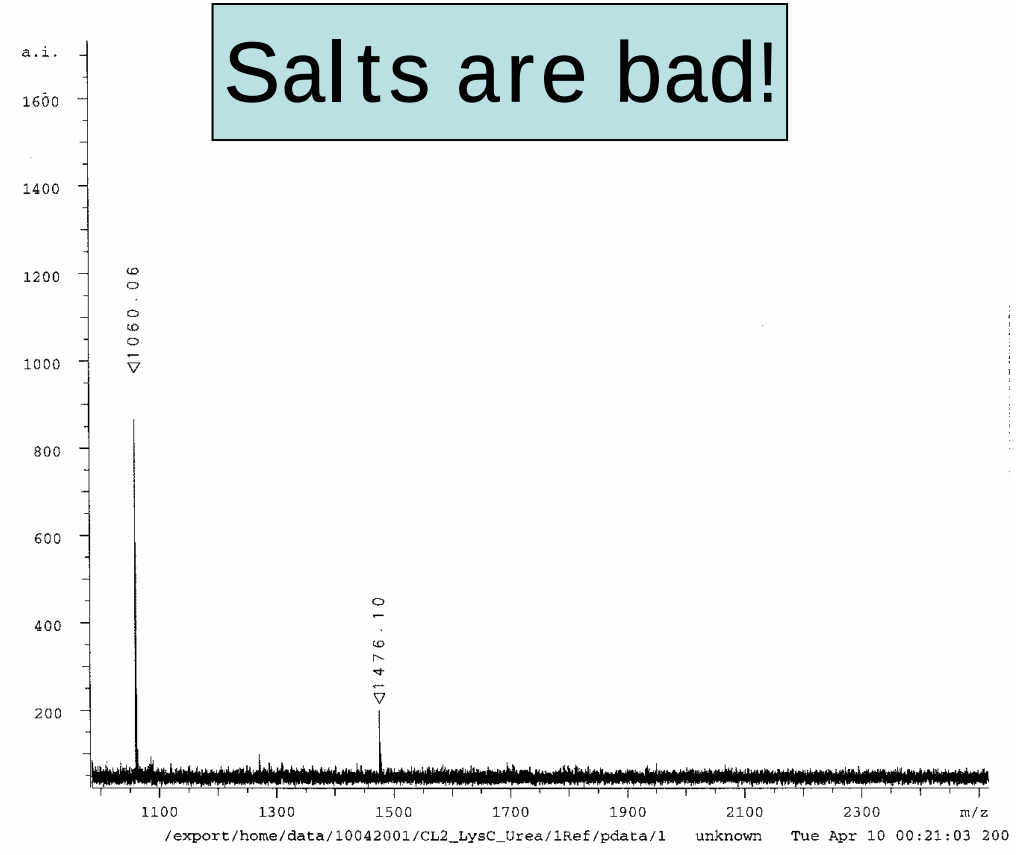
  J Am Soc Mass Spectrom 16, 471-481.

  “Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification”

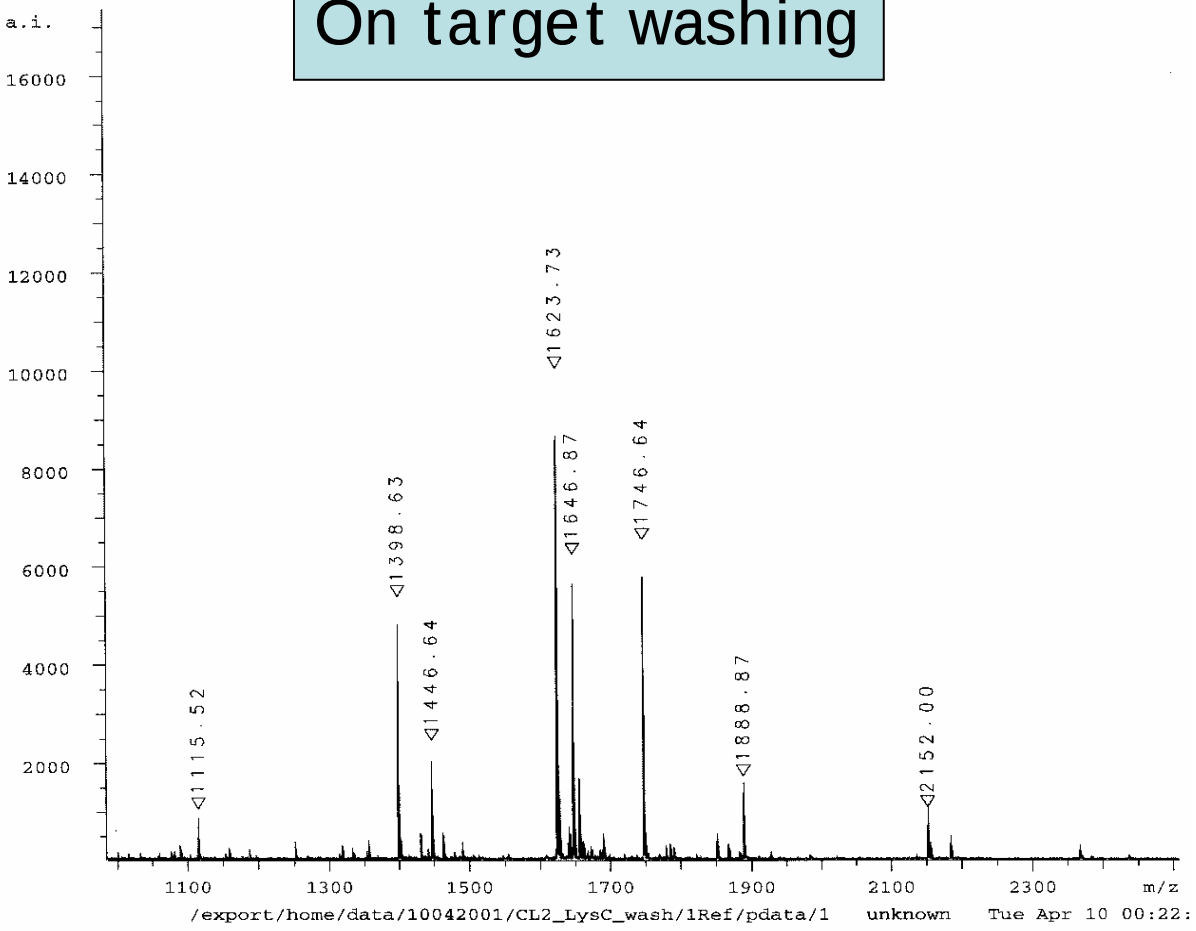
Sample preparation

It is not as the manufacturers claimed!

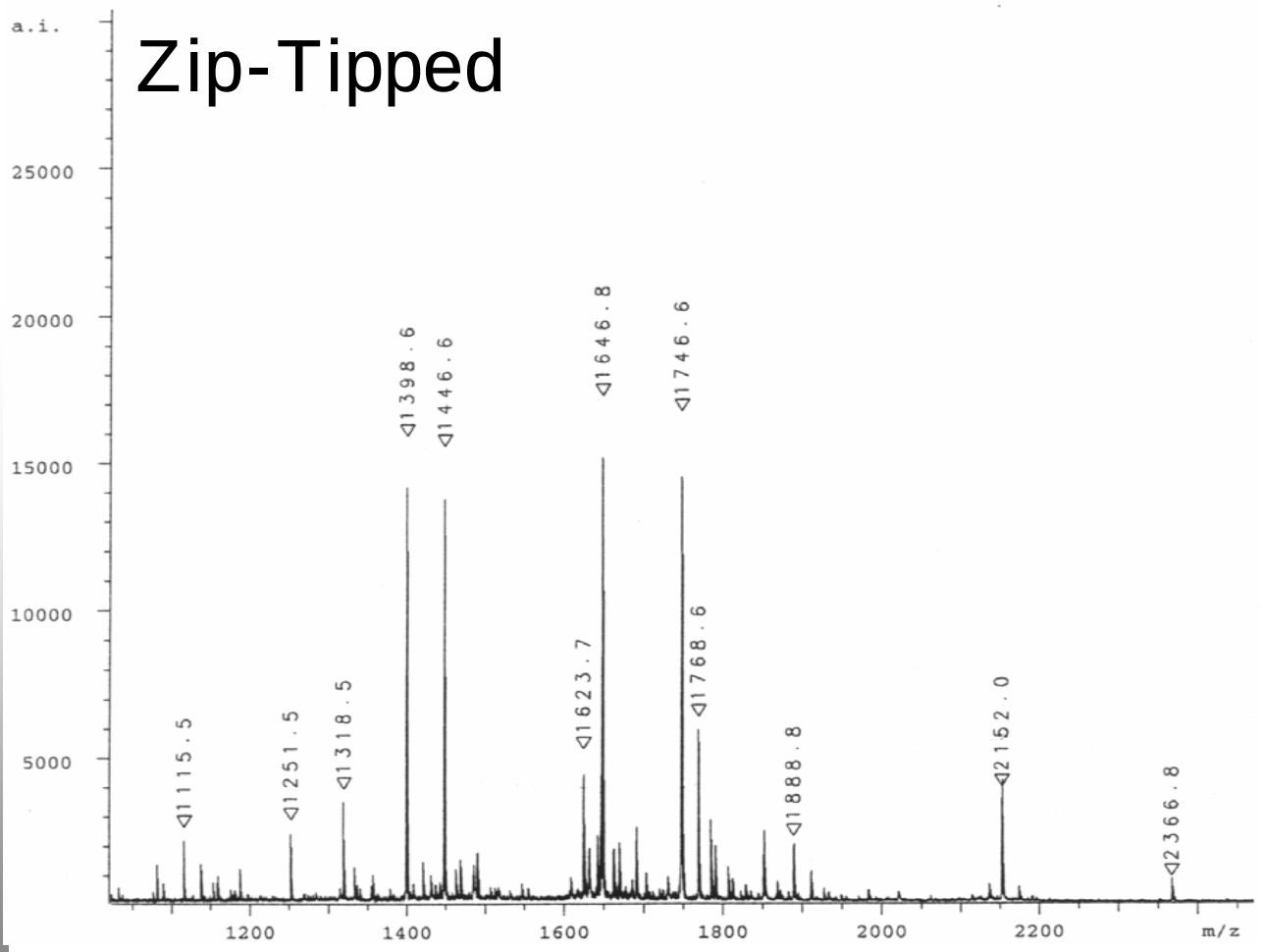
Salts are bad!



On target washing

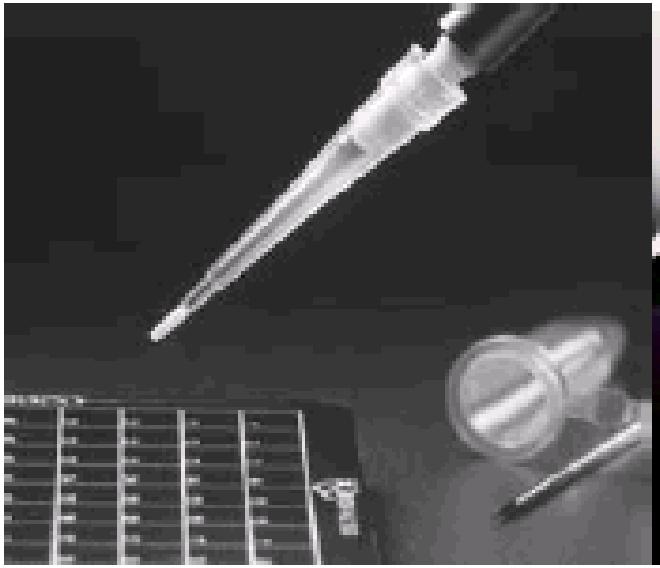


Zip-Tipped


a.i.
25000
20000
15000
10000
5000
1115.5
1251.5
1318.5
1398.6
1446.6
1623.7
1646.8
1746.6
1768.6
1888.8
2152.0
2366.8
1200
1400
1600
1800
2000
2200
m/z

This is a money game!

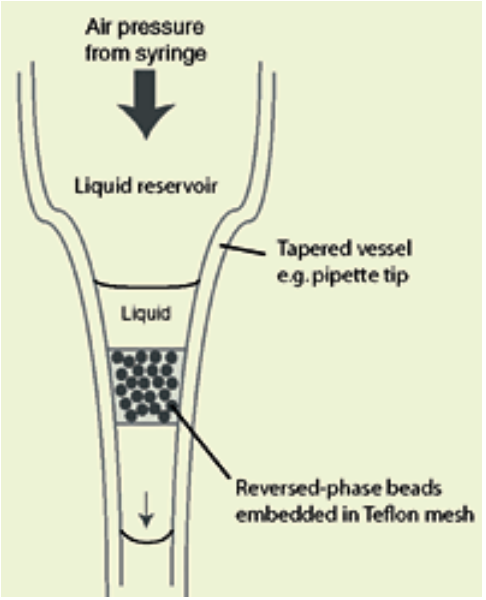
Was NT$57
Now NT$66



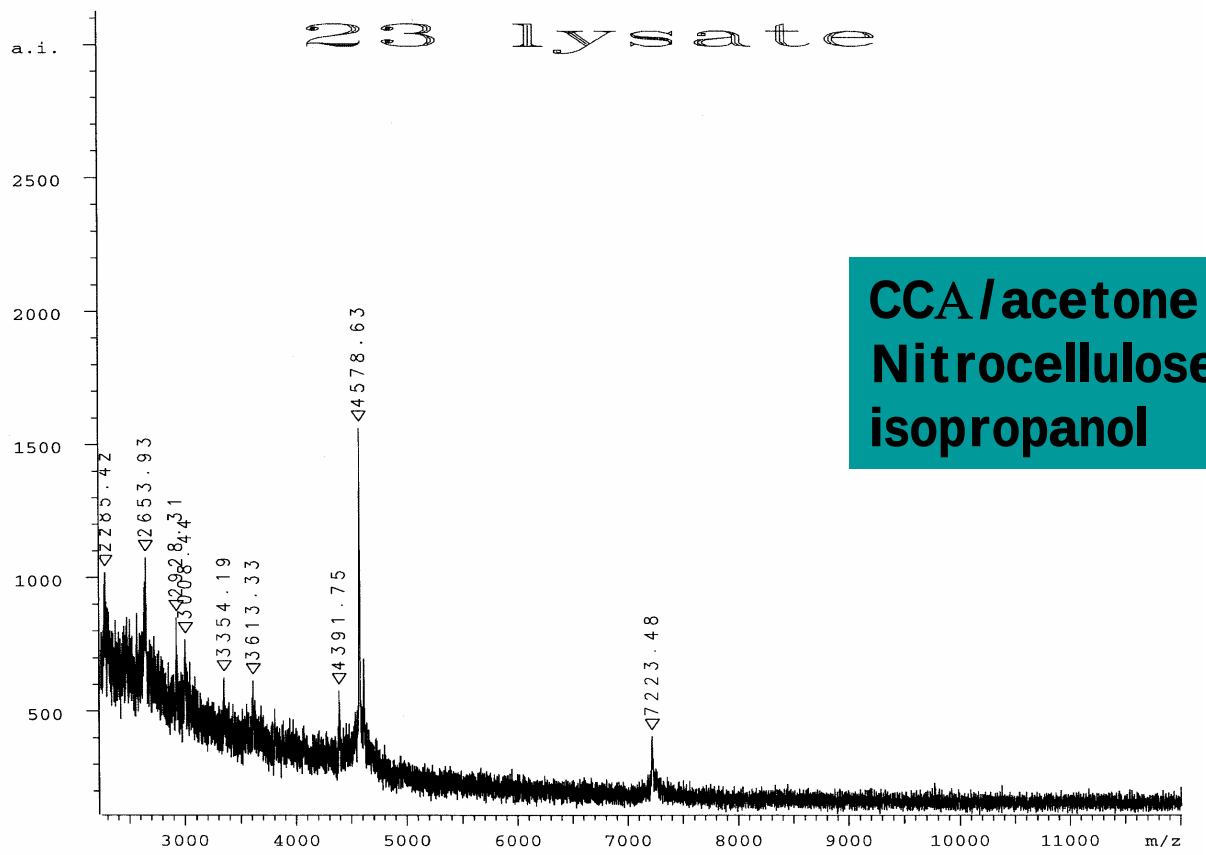
Alternative approach

Anal Chem (2003) 75, 663-666

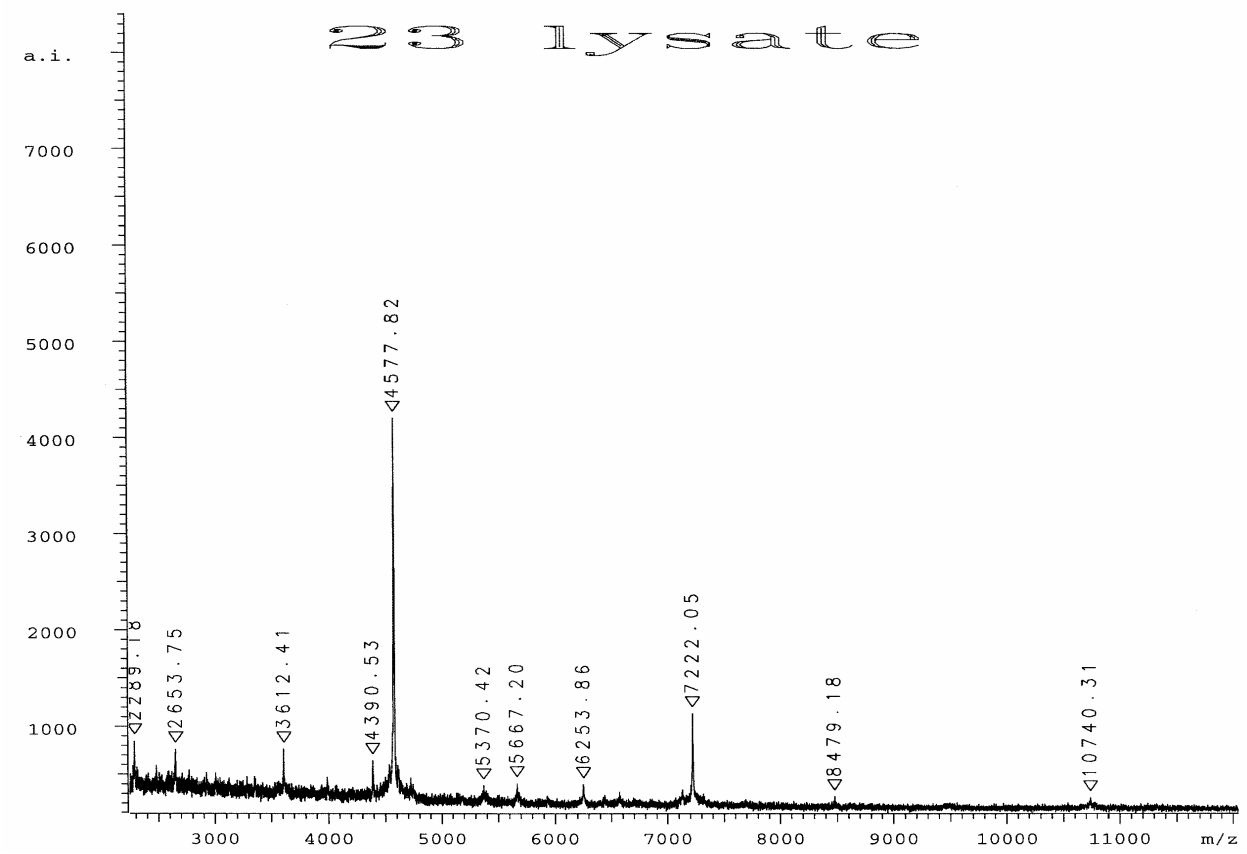
Stop and Go Extraction Tips
Or StageTips



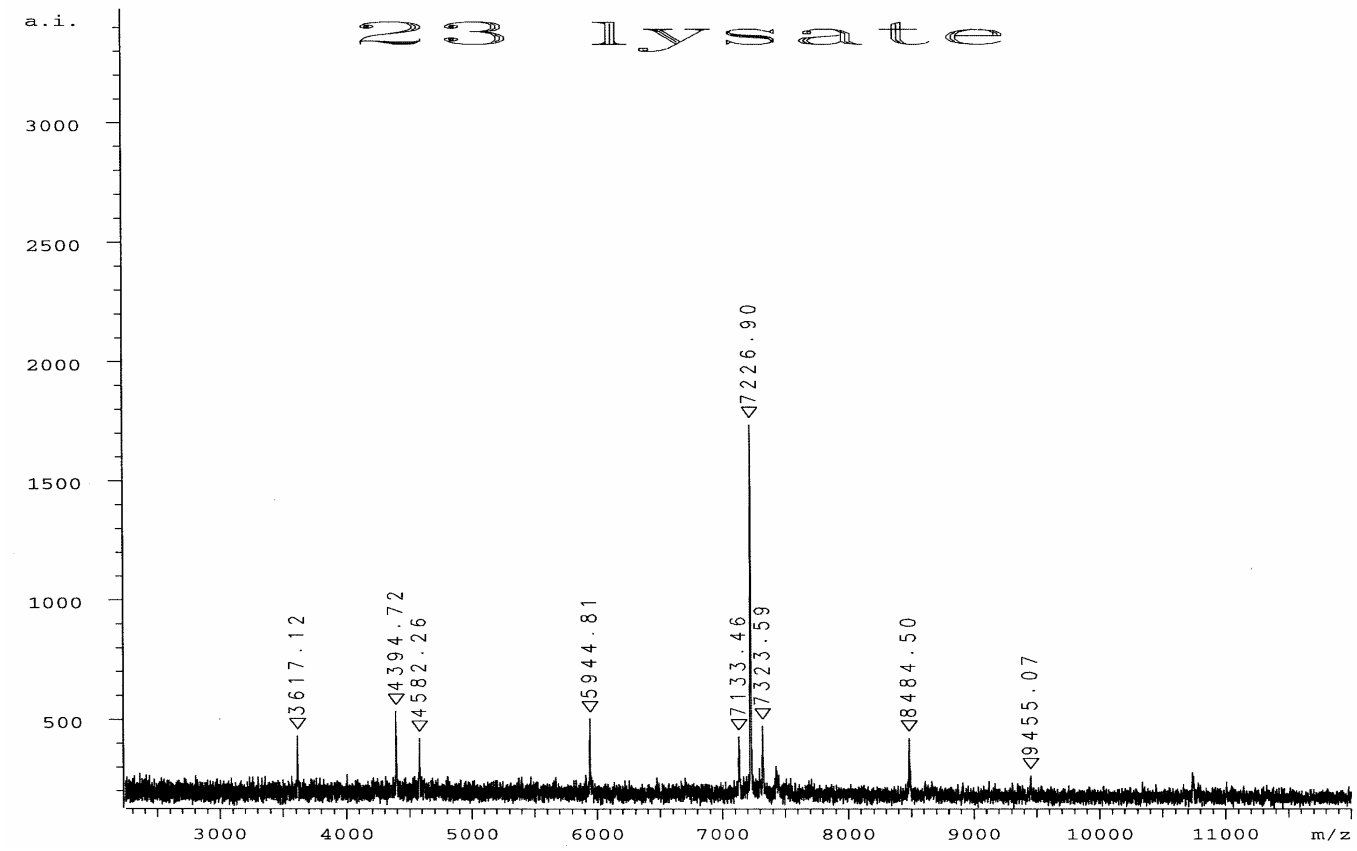
Different matrix, different results



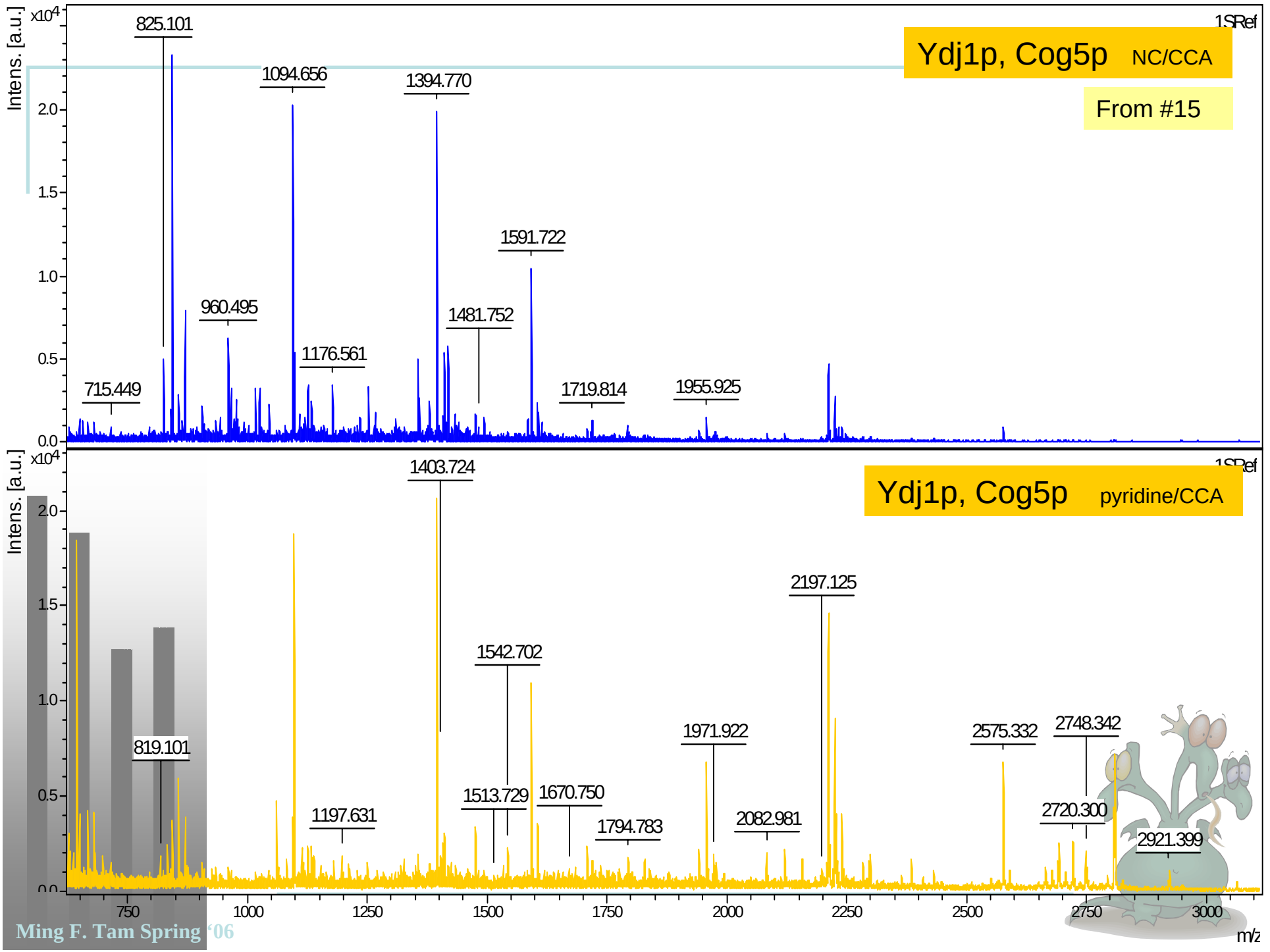
CCA/acetone
Nitrocellulose/acetone
isopropanol

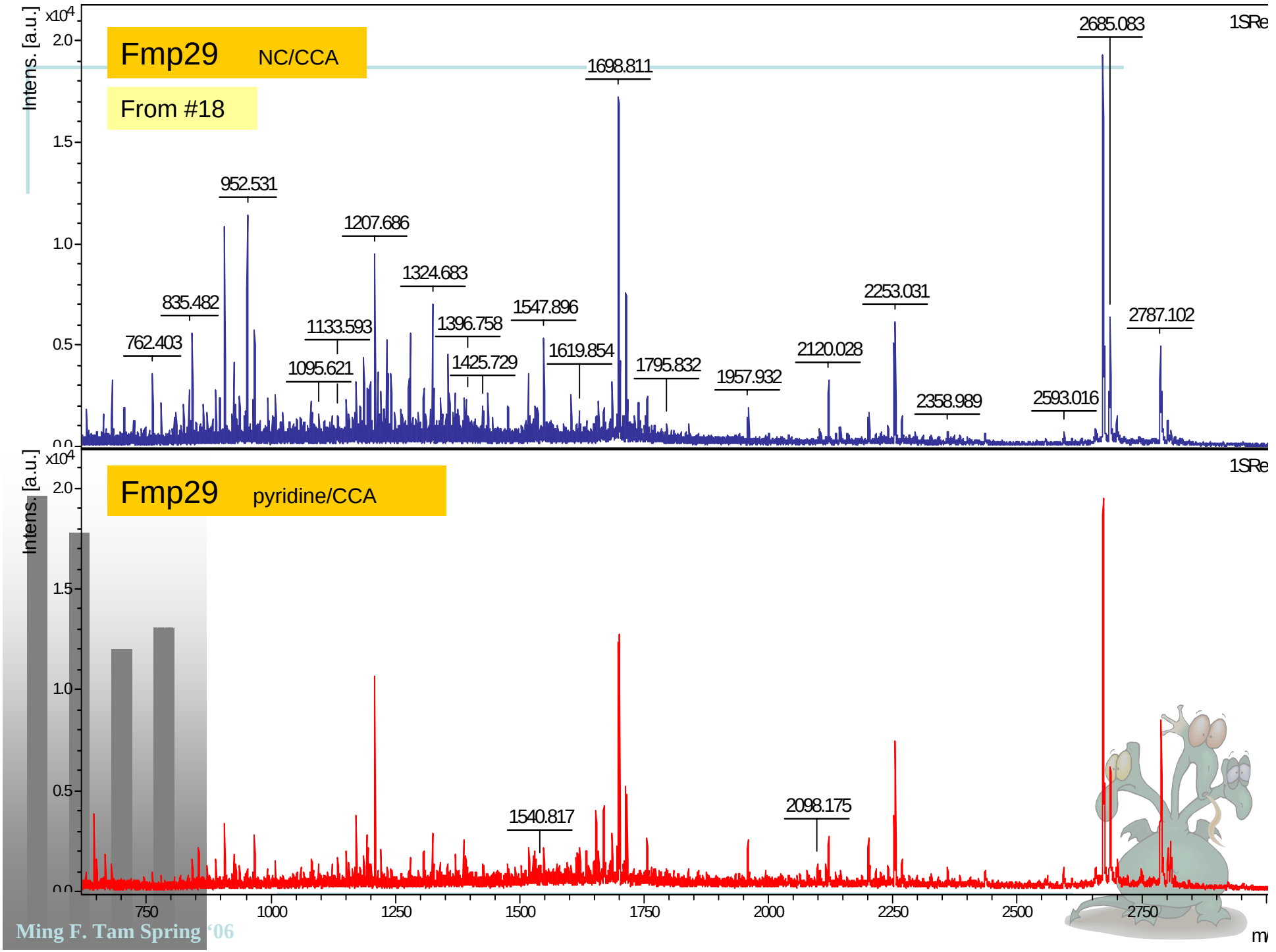


CCA/Acetonitrile/TFA



SA/Acetonitrile/TFA

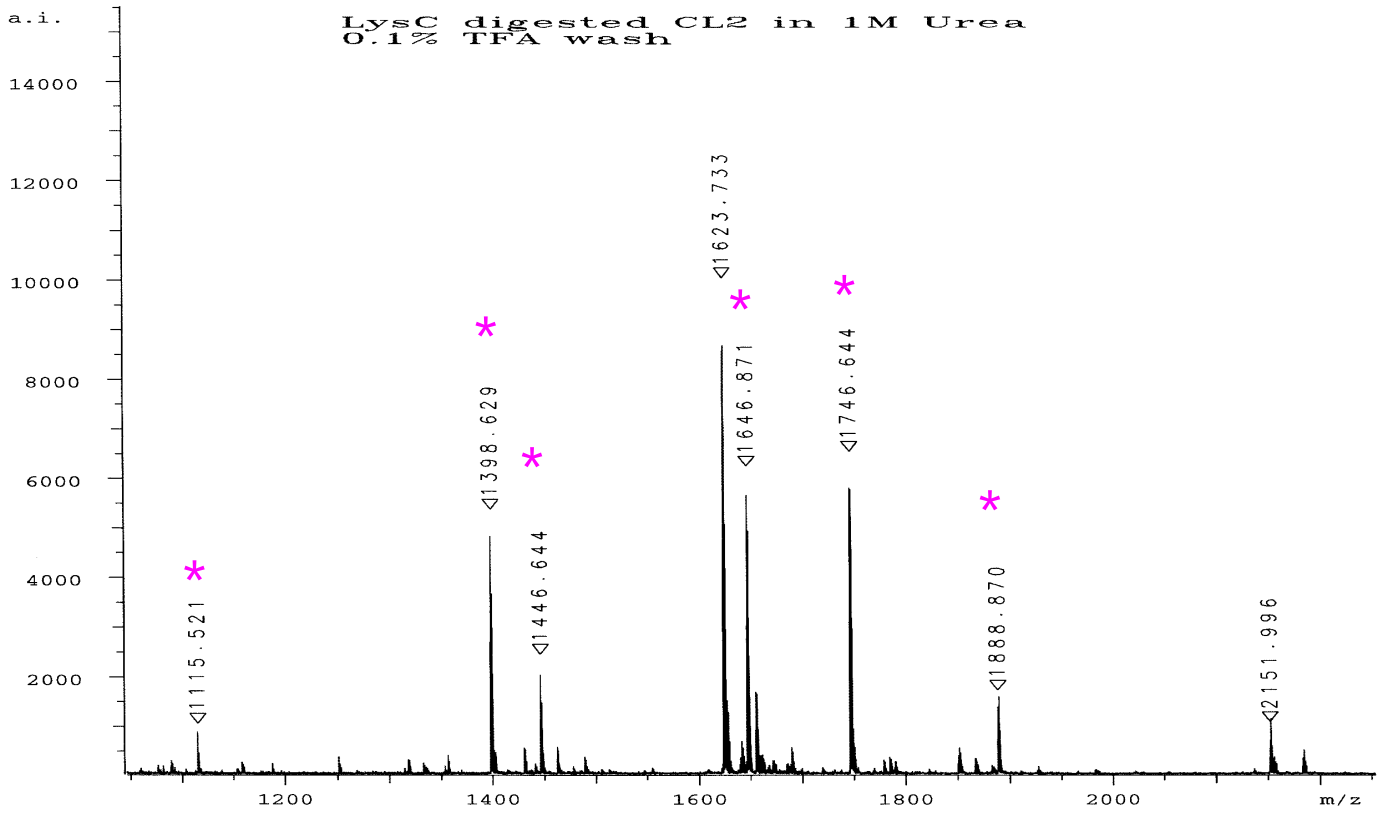

Ydj1p, Cog5p NC/CCA
From #15
Intens. [a.u.]
x10^4
825.101
1094.656
1394.770
1591.722
960.495
1481.752
1176.561
715.449
1719.814
1955.925
Ydj1p, Cog5p pyridine/CCA
1403.724
2197.125
1542.702
1971.922
2575.332
2748.342
819.101
1513.729
1670.750
1197.631
1794.783
2082.981
2720.300
2921.399
750
1000
1250
1500
1750
2000
2250
2500
2750
3000
m/z
Ming F. Tam Spring '06


Fmp29 NC/CCA
From #18
Intens. [a.u.]
x10^4
2.0
1.5
1.0
0.5
0.0
762.403
835.482
952.531
1095.621
1133.593
1207.686
1324.683
1396.758
1425.729
1547.896
1619.854
1698.811
1795.832
1957.932
2120.028
2253.031
2358.989
2593.016
2685.083
2787.102
1SRe
Fmp29 pyridine/CCA
1540.817
2098.175
750
1000
1250
1500
1750
2000
2250
2500
2750
m/
Ming F. Tam Spring '06

Data acquisition

1. Clean samples!
2. Different matrix, different results
3. Calibration—linear for unknowns beyond the range, quadratic for unknowns within range

Linear Standards: 1046.54, 2465.20
Quadratic standards: 1046.54, 1677.77, 2710.39

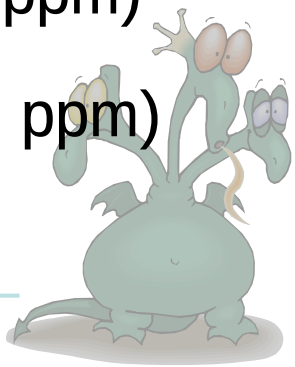


Linear vs. Quadratic

MASS

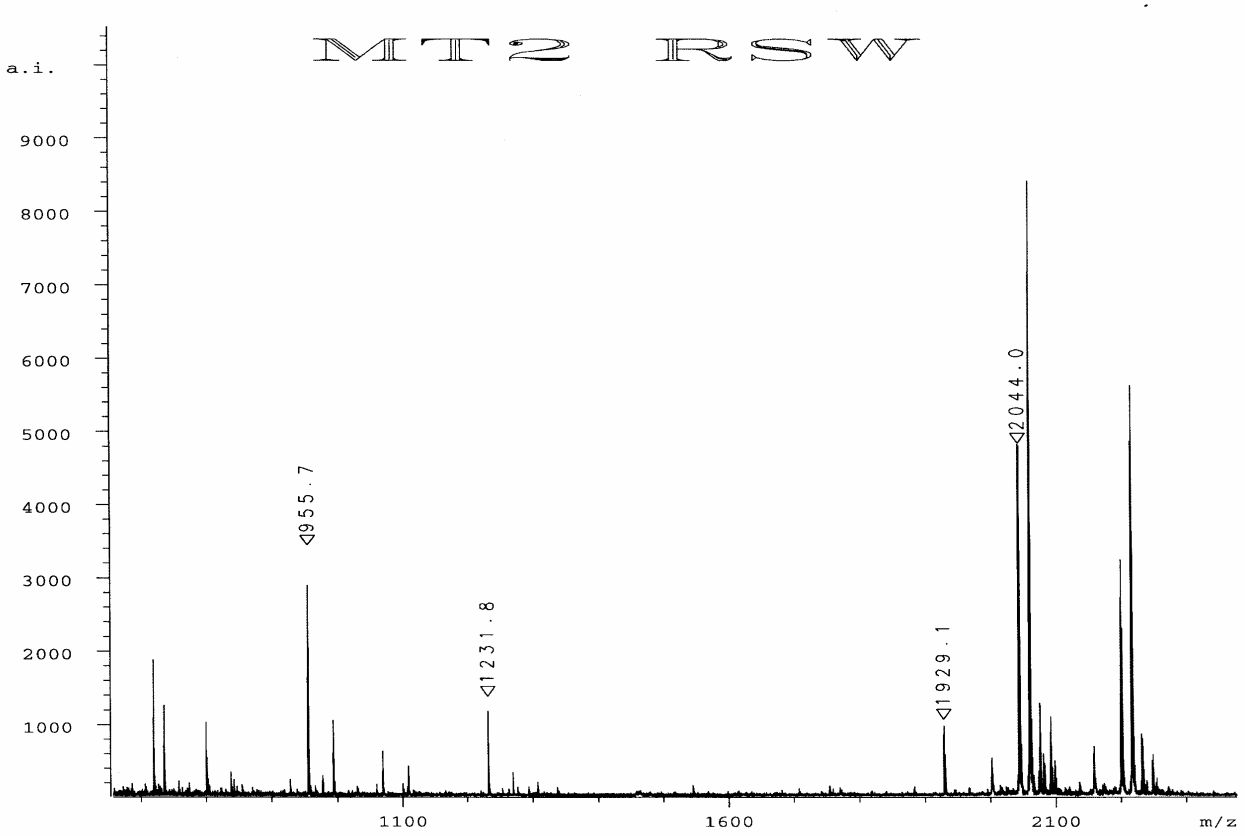
Calculated	Linear Obs.	Quadratic Obs.
1115.60	1115.47 (117 ppm)	1115.52 (71 ppm)
1398.71	1398.54 (122 ppm)	1398.63 (58 ppm)
1446.73	1446.57 (111 ppm)	1446.64 (62 ppm)
1646.96	1646.78 (109 ppm)	1646.87 (55 ppm)
1746.75	1746.53 (126 ppm)	1746.64 (63 ppm)
1889.00	1889.76 (402 ppm)	1888.87 (69 ppm)

Linear Standards: 1046.54, 2465.20
Quadratic standards: 1046.54, 1677.77, 2710.39



But internal standards are much better!

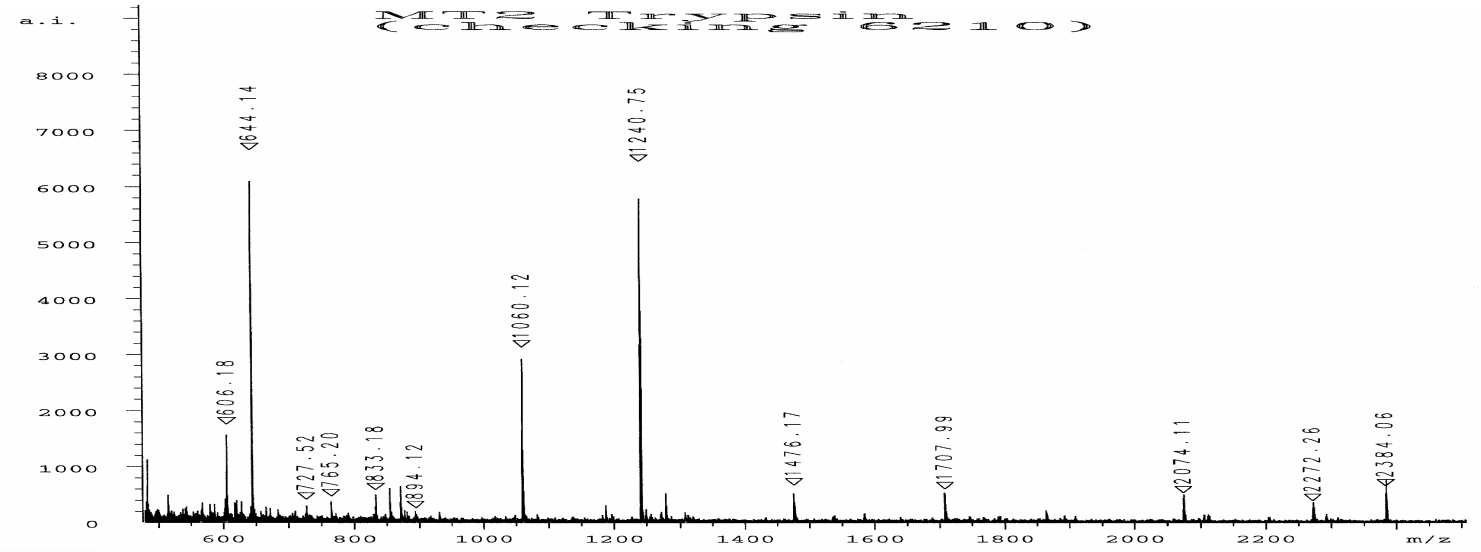
379.09, 1046.54, 2710.39

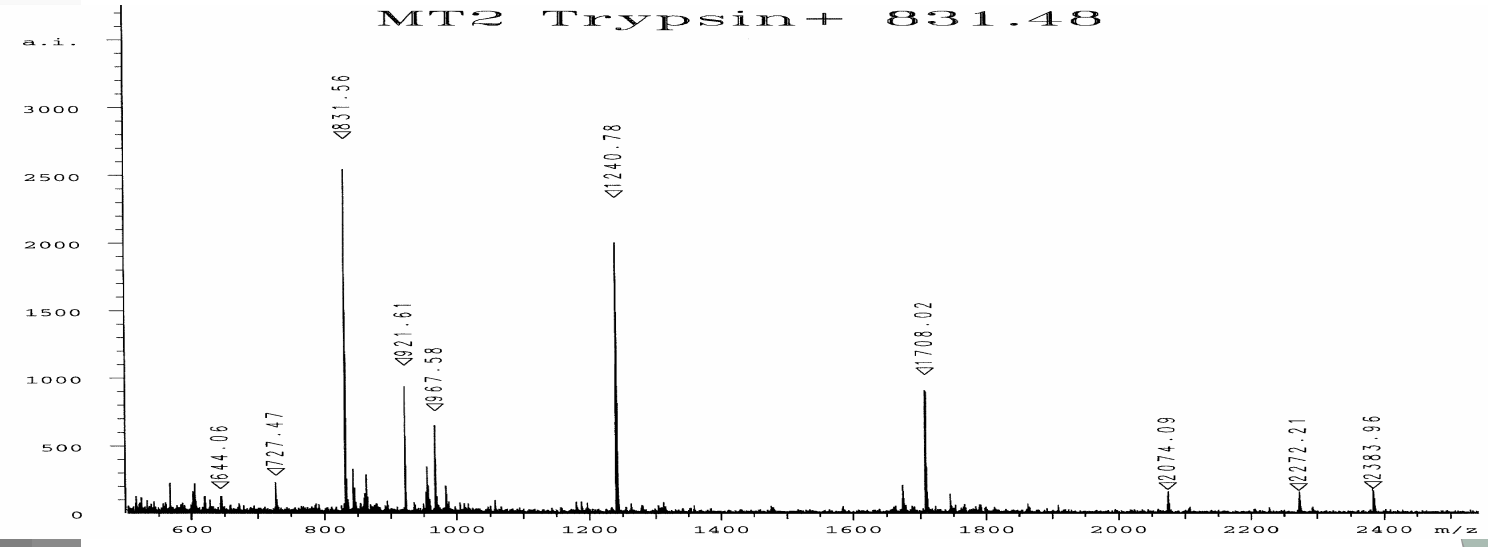


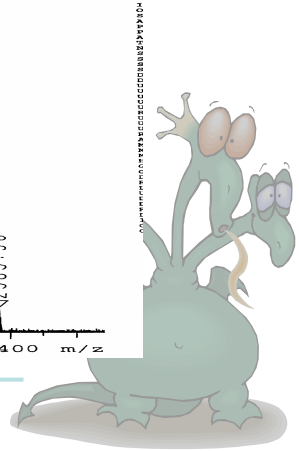
External vs. Internal

Real	Ext. Obs.	Int. Obs.
955.57	955.70 (136 ppm)	955.58 (10 ppm)
1231.69	1231.79 (81 ppm)	1231.70 (8 ppm)
1928.94	1929.10 (83 ppm)	1929.00 (31 ppm)
2043.87	2044.03 (78 ppm)	2043.92 (24 ppm)

But watch out for Signal suppression!!!

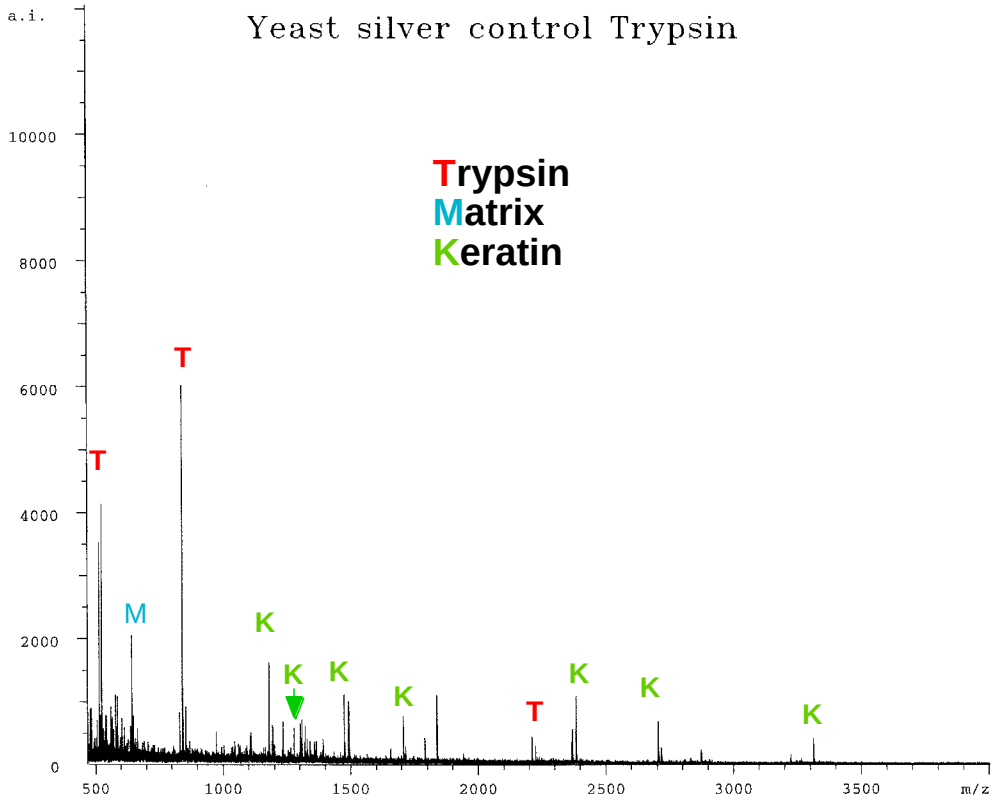


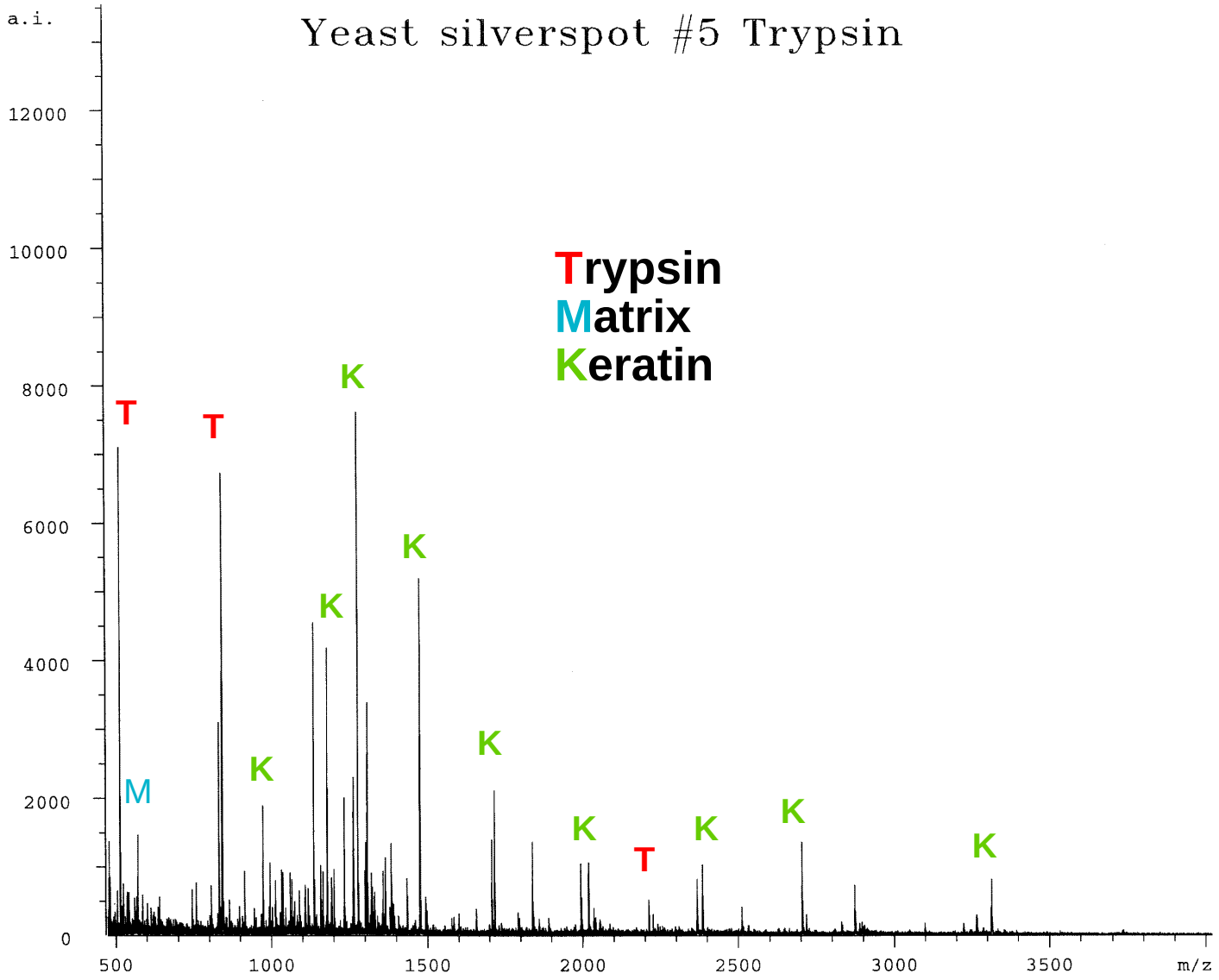


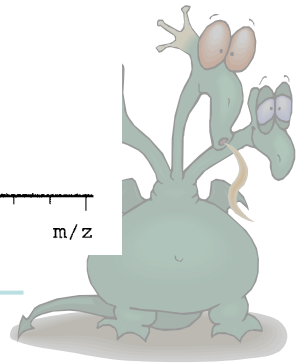


You can never be too clean!

MALDI spectrum of a silver stained empty gel spot




Yeast silverspot #5 Trypsin
a.i.
12000
10000
8000
6000
4000
2000
0
500
1000
1500
2000
2500
3000
3500
m/z
Trypsin
Matrix
Keratin
T
T
K
K
K
M
K
K
K
T
K
K
K



Human Keratins

K1 (skin) (M_r): **65847**; pI: **8.16** K2E (dandruff) (M_r): **65825**; pI: **8.07**

K9 (skin) (M_r): **61950**; pI: **5.14** K10 (dandruff) (M_r): **59492**; pI: **5.17**

MH^+

897.41	K9	1037.52	K2E	1060.56	K9
1066.49	K9	1107.54		1165.58	K10
1193.61	K2E	1320.58	K2E	1365.63	K10
1381.64	K10	1383.68	K1	1475.74	K1
1707.76	K10	1716.85	K1	1838.91	K2E
1993.97	K1	2025.94	K10	2367.26	K10
2383.94	K1	2510.12	K9	2705.15	K9
2807.30		2831.19	K2E	3312.30	K1

http://prospector.ucsf.edu/ucsfhtml4.0/misc/**trypsin.**htm

Porcine trypsin

Entries in italics are for the variant protein I_{20} -> V:

From	To	MH^+	Sequence
52	53	262.14	SR
54	57	515.32	IQVR
108	115	842.50	VATVSLPR
209	216	906.50	NKPGVYTK
148	157	1006.48	APVLSDSSCK
98	107	1045.56	LSSPATLNSR
134	147	1469.72	SSGSSYPSLLQCLK
217	231	1736.84	VCNYVNWIQQTIAAN
116	133	1768.79	SCAAAGTECLISGWGNTK
158	178	2158.02	SSYPGQITGNMICVGFLEGGK
58	77	2211.10	LGEHNIDVLEGNEQFINAAK
78	97	2283.17	IITHPNFNGNTLDNDIMLIK
179	208	3013.32	DSCQGDSGG...SWGYGCAQK
9	*51*	*4475.09*	*IVGGYTCAA...VVSAAHCYK*
9	51	4489.11	IVGGYTCAA...VVSAAHCYK

Alternatives to 2-D gel

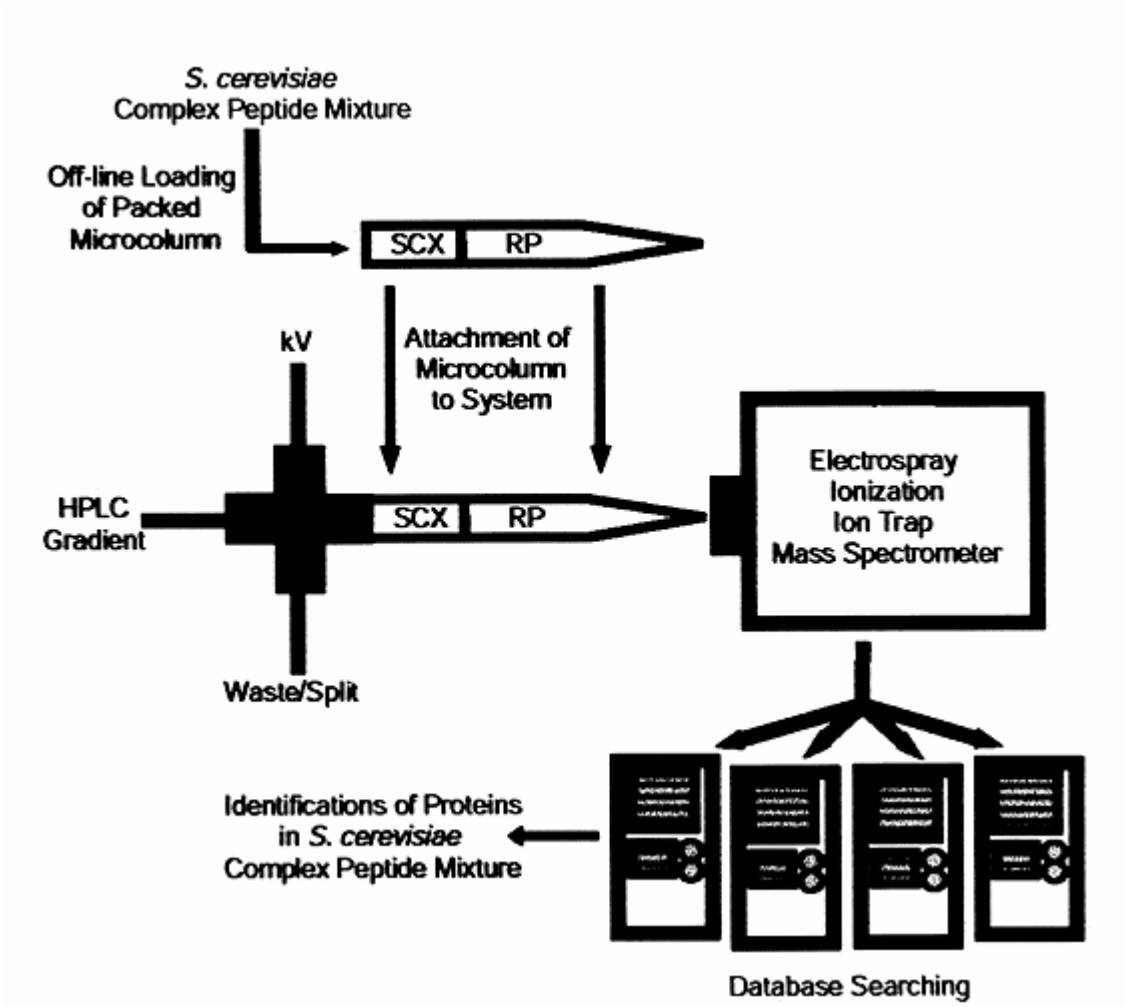
Multidimensional protein identification technology

LC-LC/MS-MS (MUDPIT) to replace Gel/MS-MS

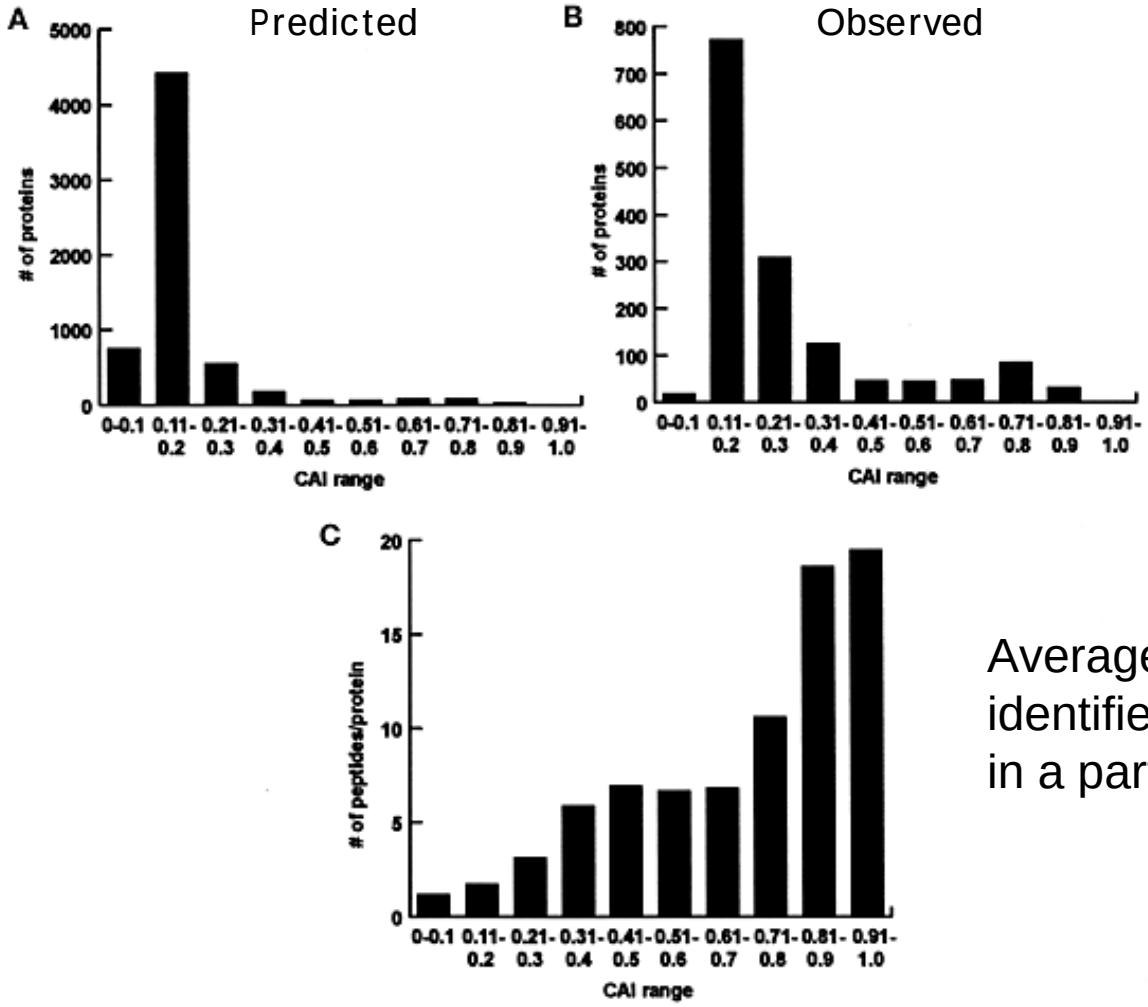
Washburn et al. *Nature Biotech* 2001, 19: 242-247
"Large scale analysis of the yeast proteome by MultiDimensional Protein Identification Technology"

Peptide digests separated with cation exchanger/ reverse-phase on line with an ESI/ion trap.

5 micron particles, 0.1 mm ID, 4 and 10 cm long

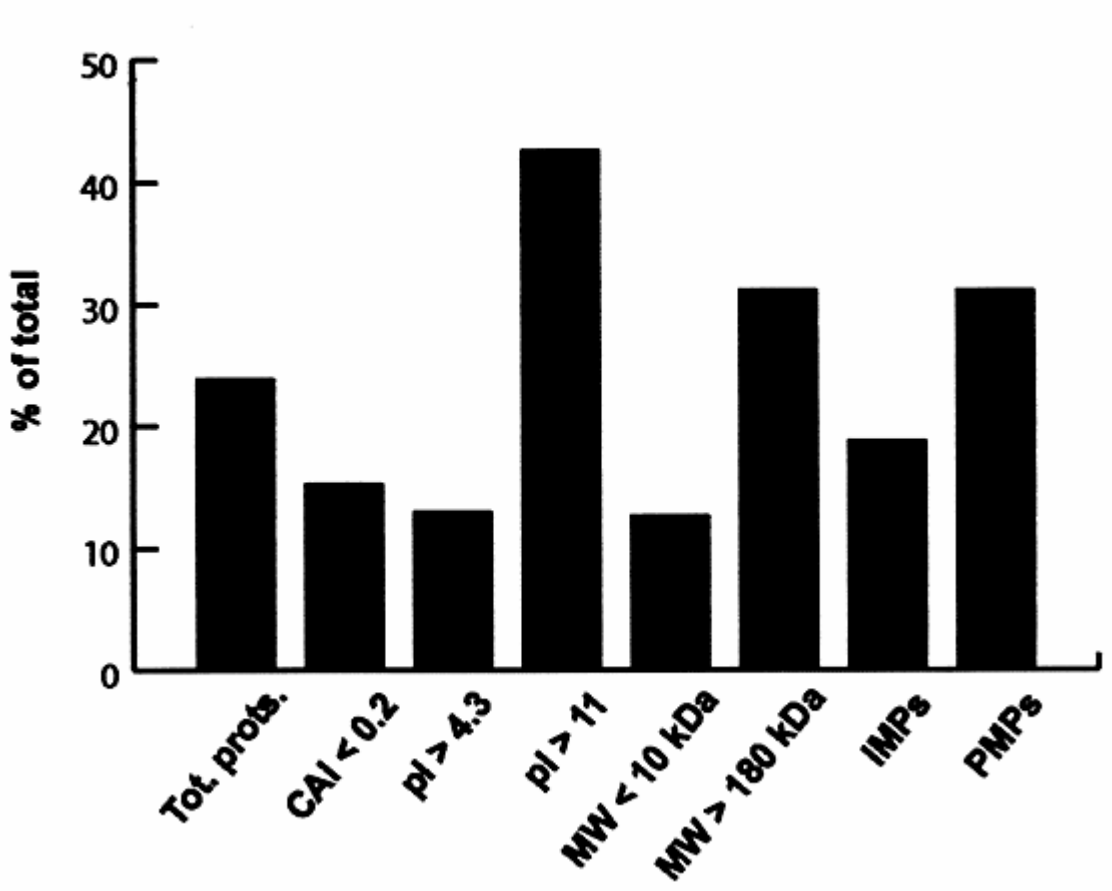


6212 ORF, 1484 proteins identified



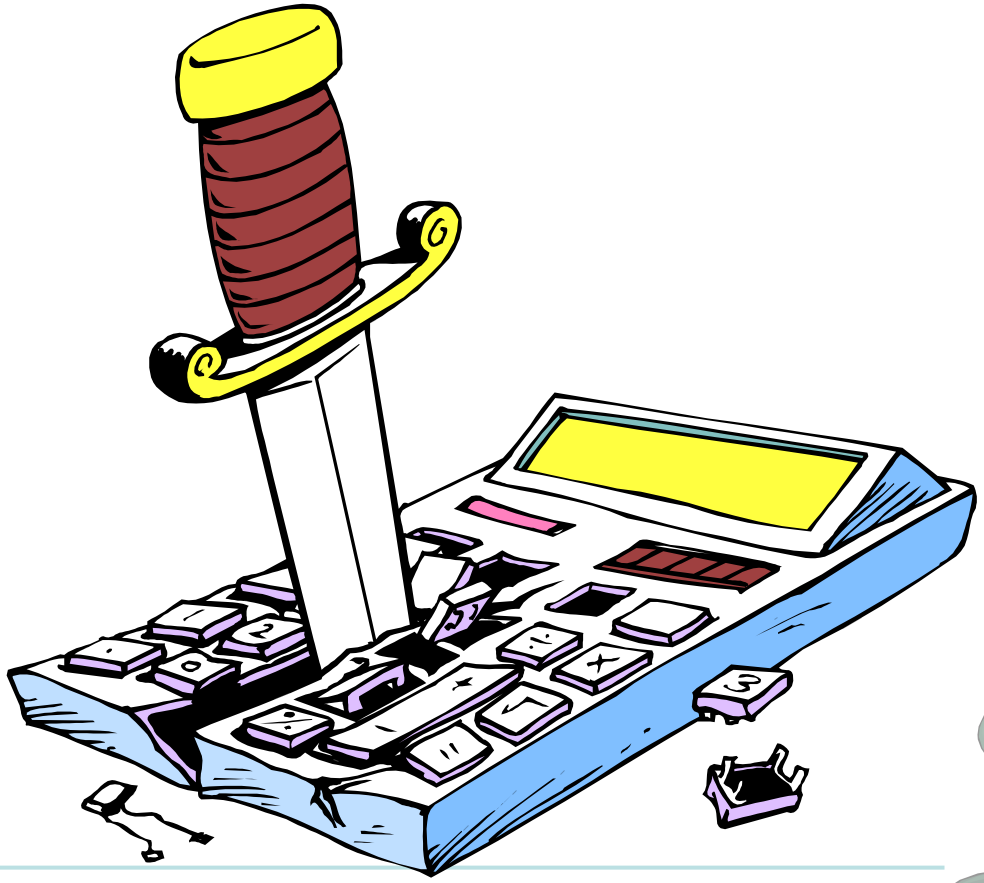
Average # of peptides identified for each protein in a particular CAI range

Sensitivity of MudPIT to a wide variety of protein classes



I- and PRP = integral and peripheral membrane protein

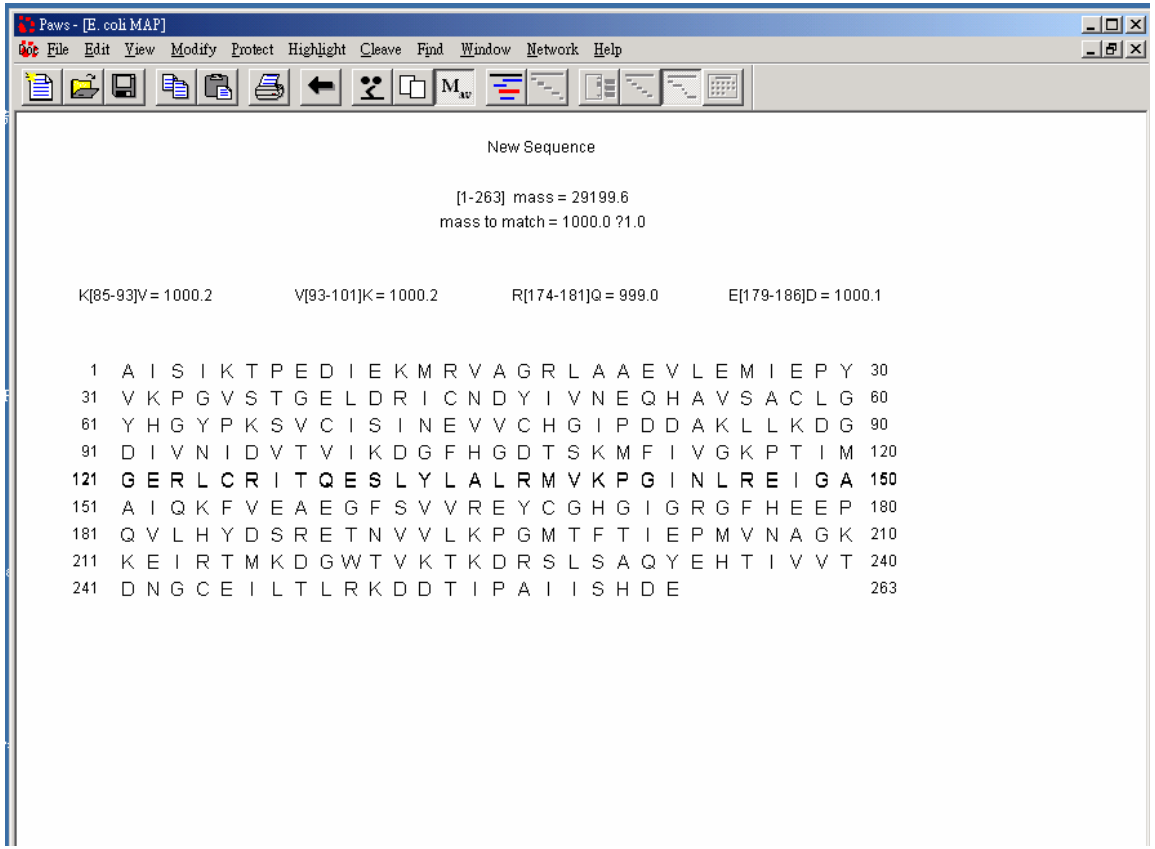
Search the Databases



General free programs

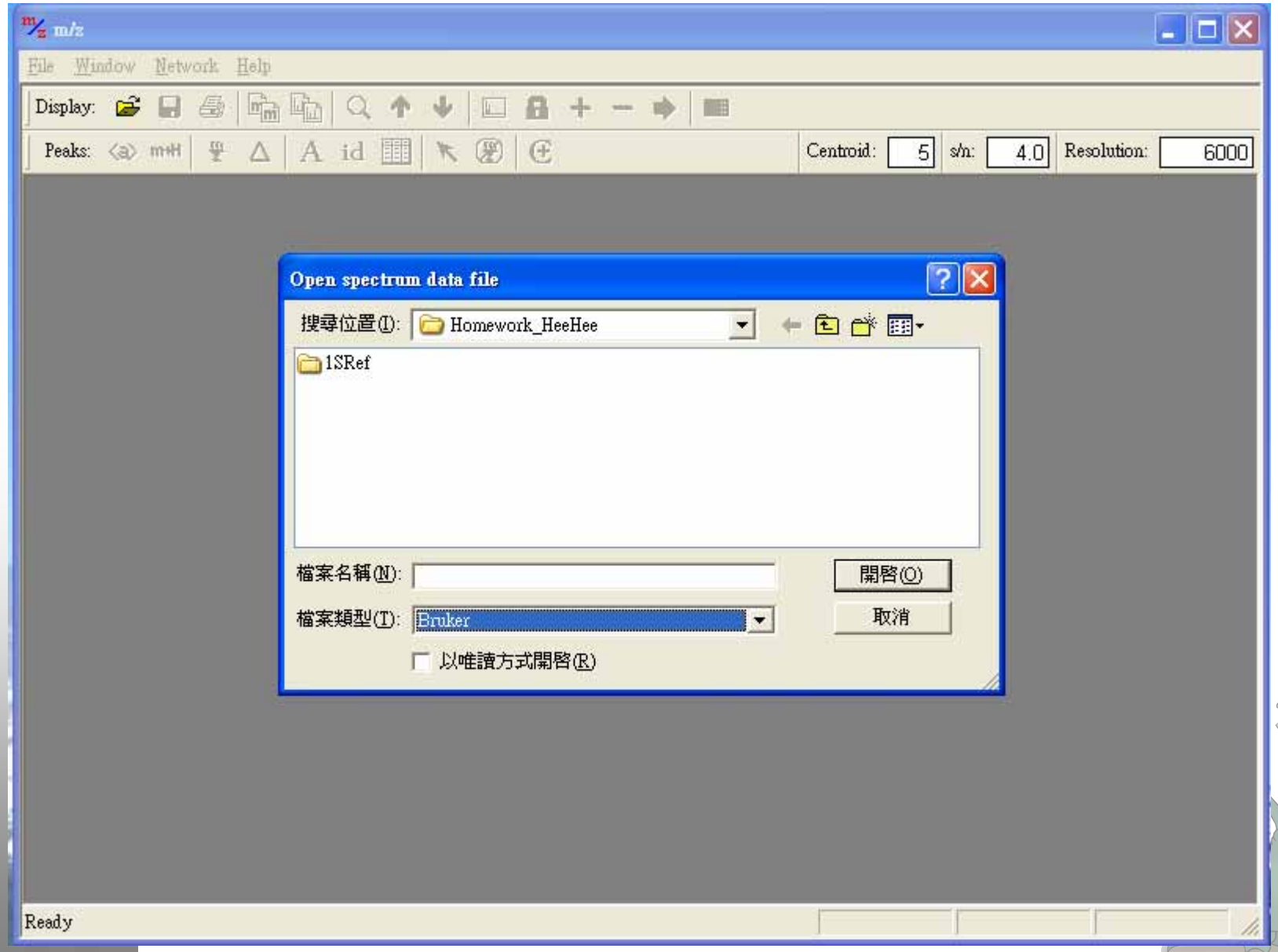
PAWS http://65.219.84.5/paws.html

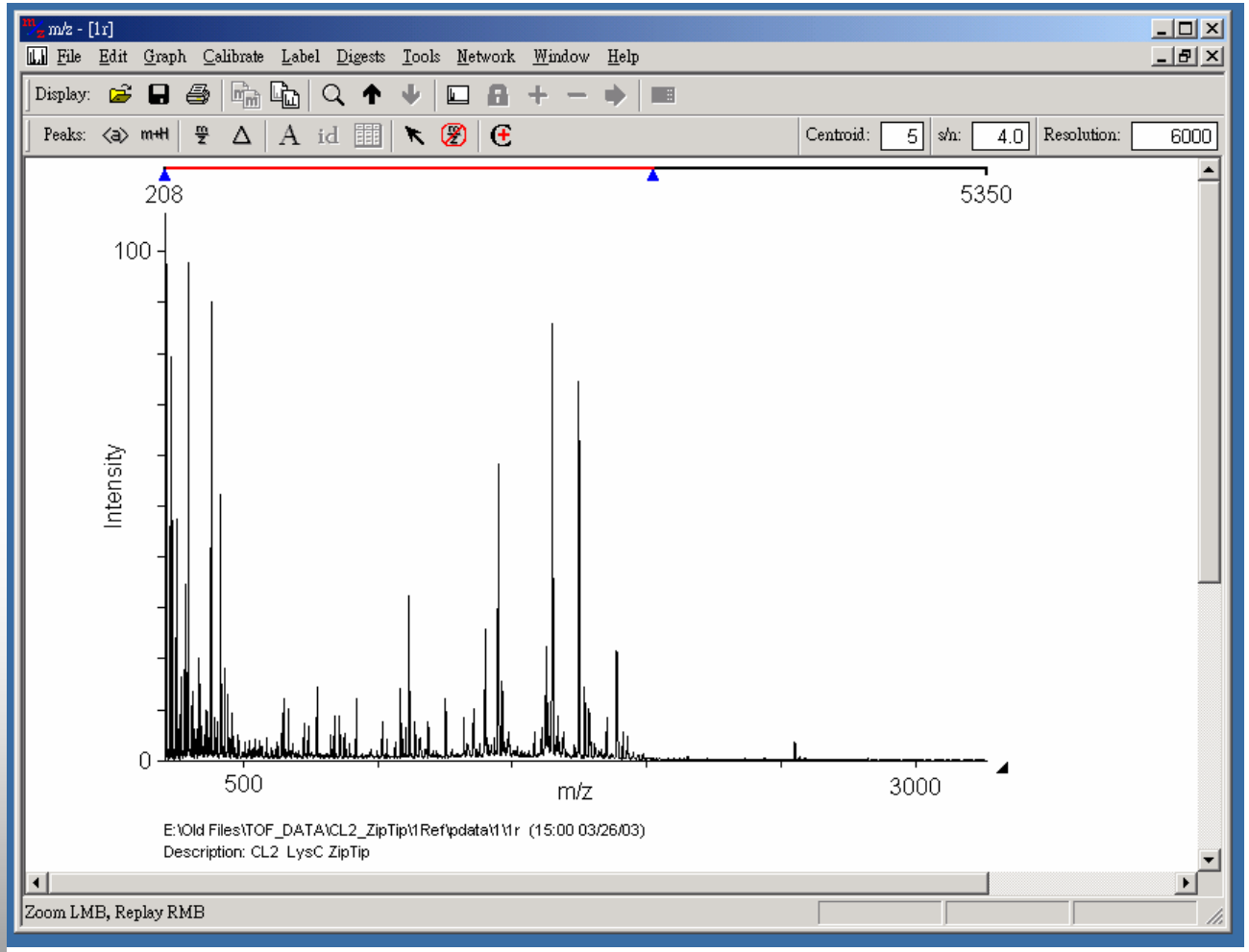
From Rockefeller U.→ Genomic Solutions



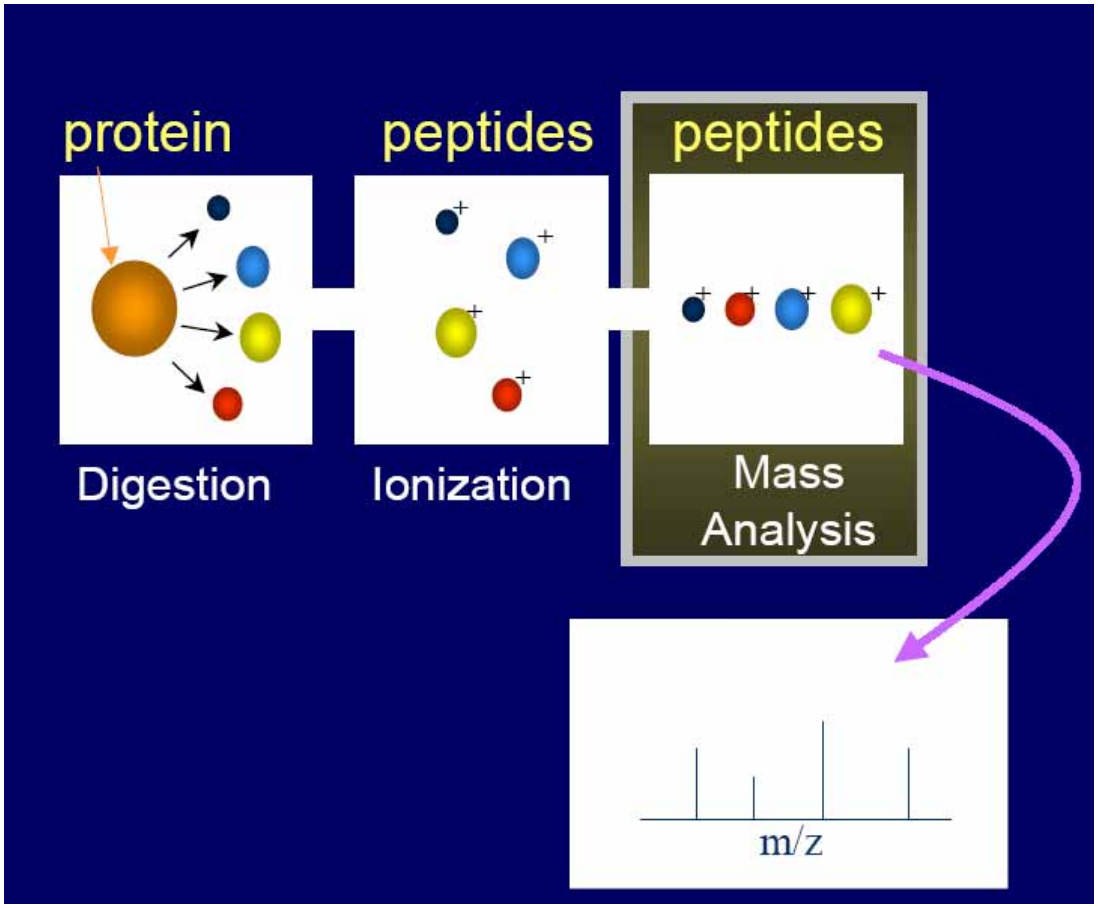
Moverz http://65.219.84.5/moverz.html

http://bioinformatics.genomicsolutions.com/MoverZDL.html



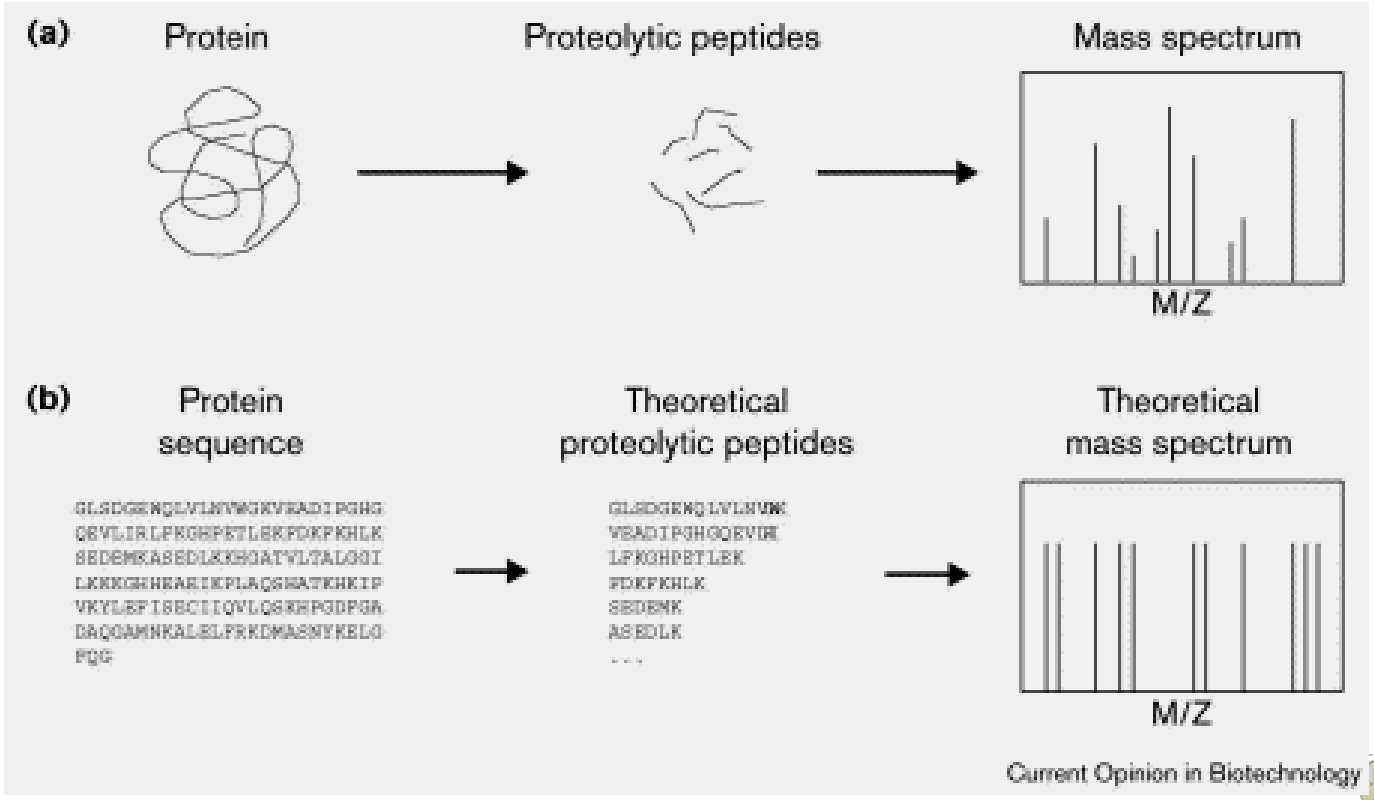

m/z - [1r]
File Edit Graph Calibrate Label Digests Tools Network Window Help
Display:
Peaks:
Centroid: 5
s/n: 4.0
Resolution: 6000
208
5350
100
0
Intensity
500
m/z
3000
E:\Old Files\TOF_DATA\CL2_ZipTip\1Ref\pdata\1\1r (15:00 03/26/03)
Description: CL2 LysC ZipTip
Zoom LMB, Replay RMB

Two types of data:

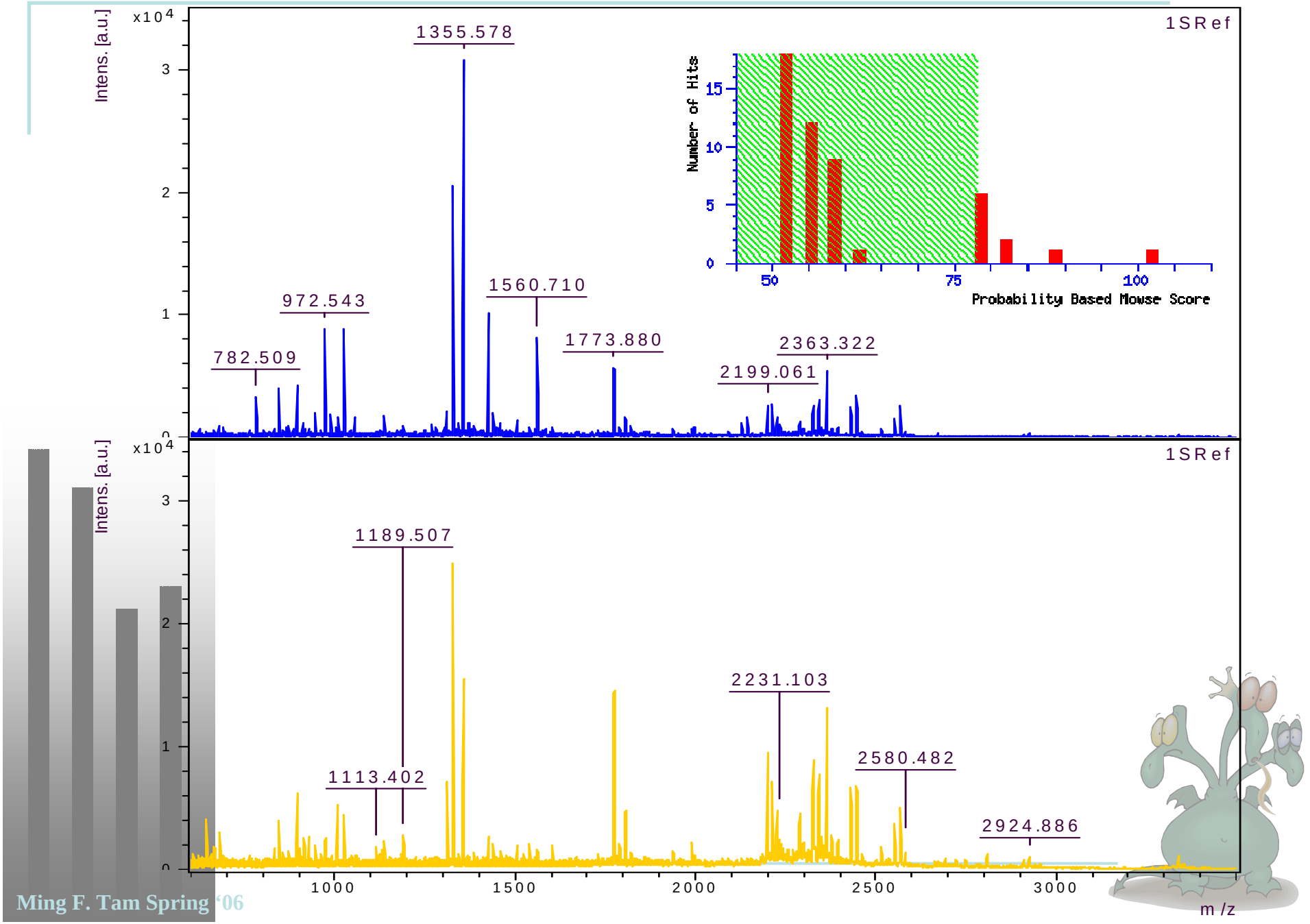


Peptide mapping/peptide mass fingerprinting

Peptide mapping



Sbp1/R1-3


Intens. [a.u.]
x10 4
1SRef
1355.578
1560.710
972.543
1773.880
782.509
2363.322
2199.061
Number of Hits
Probability Based Mowse Score
1189.507
1113.402
2231.103
2580.482
2924.886
m /z
Ming F. Tam Spring '06

10	20	30	40	50	60
MGKEKSHINV	VVIGHVDSGK	STTTGHLIYK	CGGIDKRTIE	KFEKEAAELG	KGSFKYAWVL
70	80	90	100	110	120
DKLKAERERG	ITIDIALWKF	ETPKYQVTVI	DAPGHRDFIK	NMITGTSQAD	CAILIIAGGV
130	140	150	160	170	180
GEFEAGISKD	GQTREHALLA	FTLGVRQLIV	AVNKMDSVKW	DESRFQEIVK	ETSNFIKKVG
190	200	210	220	230	240
YNPKTVPFVP	ISGWNGDNMI	EATTNAPWYK	GWEKETKAGV	VKGKTLLEAI	DAIEQPSRPT
250	260	270	280	290	300
DKPLRLPLQD	VYKIGGIGTV	PVGRVETGVI	KPGMVVTFAP	AGVTTEVKSV	EMHHEQLEQG
310	320	330	340	350	360
VPGDNVGFNV	KNVSVKEIRR	GNVCGDAKND	PPKGCASFNA	TVIVLNHPGQ	ISAGYSPVLD
370	380	390	400	410	420
CHTAHIACRF	DELLEKNDRR	SGKKLEDHPK	FLKSGDAALV	KFVPSKPMCV	EAFSEYPPLG
430	440	450	460		
RFAVRDMRQT	VAVGVIKSVD	KTEKAAKVTK	AAQKAAKK		

Sbp1/R1—Lane 3
TEF1

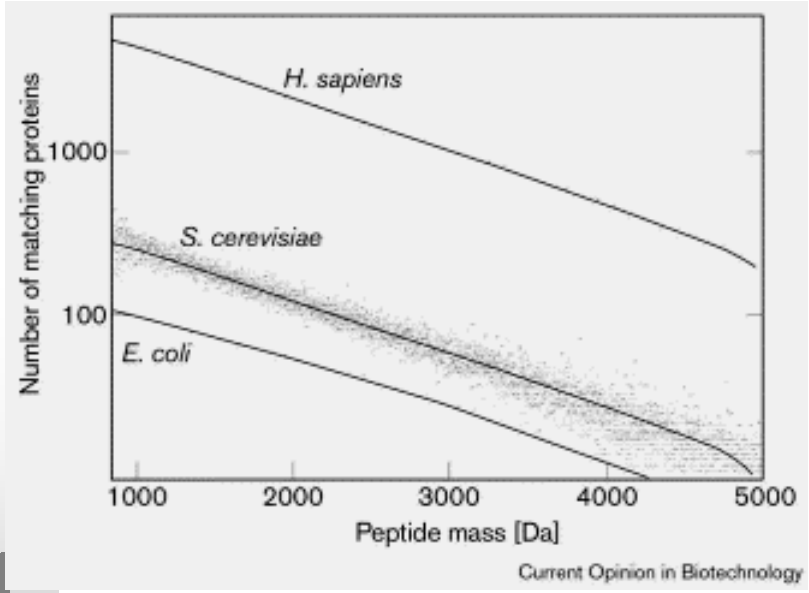
34.1% coverage

10	20	30	40	50	60
MGKEKSHINV	VVIGHVDSGK	STTTGHLIYK	CGGIDKRTIE	KFEKEAAELG	KGSFKYAWVL
70	80	90	100	110	120
DKLKAERERG	ITIDIALWKF	ETPKYQVTVI	DAPGHRDFIK	NMITGTSQAD	CAILIIAGGV
130	140	150	160	170	180
GEFEAGISKD	GQTREHALLA	FTLGVRQLIV	AVNKMDSVKW	DESRFQEIVK	ETSNFIKKVG
190	200	210	220	230	240
YNPKTVPFVP	ISGWNGDNMI	EATTNAPWYK	GWEKETKAGV	VKGKTLLEAI	DAIEQPSRPT
250	260	270	280	290	300
DKPLRLPLQD	VYKIGGIGTV	PVGRVETGVI	KPGMVVTFAP	AGVTTEVKSV	EMHHEQLEQG
310	320	330	340	350	360
VPGDNVGFNV	KNVSVKEIRR	GNVCGDAKND	PPKGCASFNA	TVIVLNHPGQ	ISAGYSPVLD
370	380	390	400	410	420
CHTAHIACRF	DELLEKNDRR	SGKKLEDHPK	FLKSGDAALV	KFVPSKPMCV	EAFSEYPPLG
430	440	450	460		
RFAVRDMRQT	VAVGVIKSVD	KTEKAAKVTK	AAQKAAKK		

Sbp1—Lane 3
TEF1

37.6% coverage

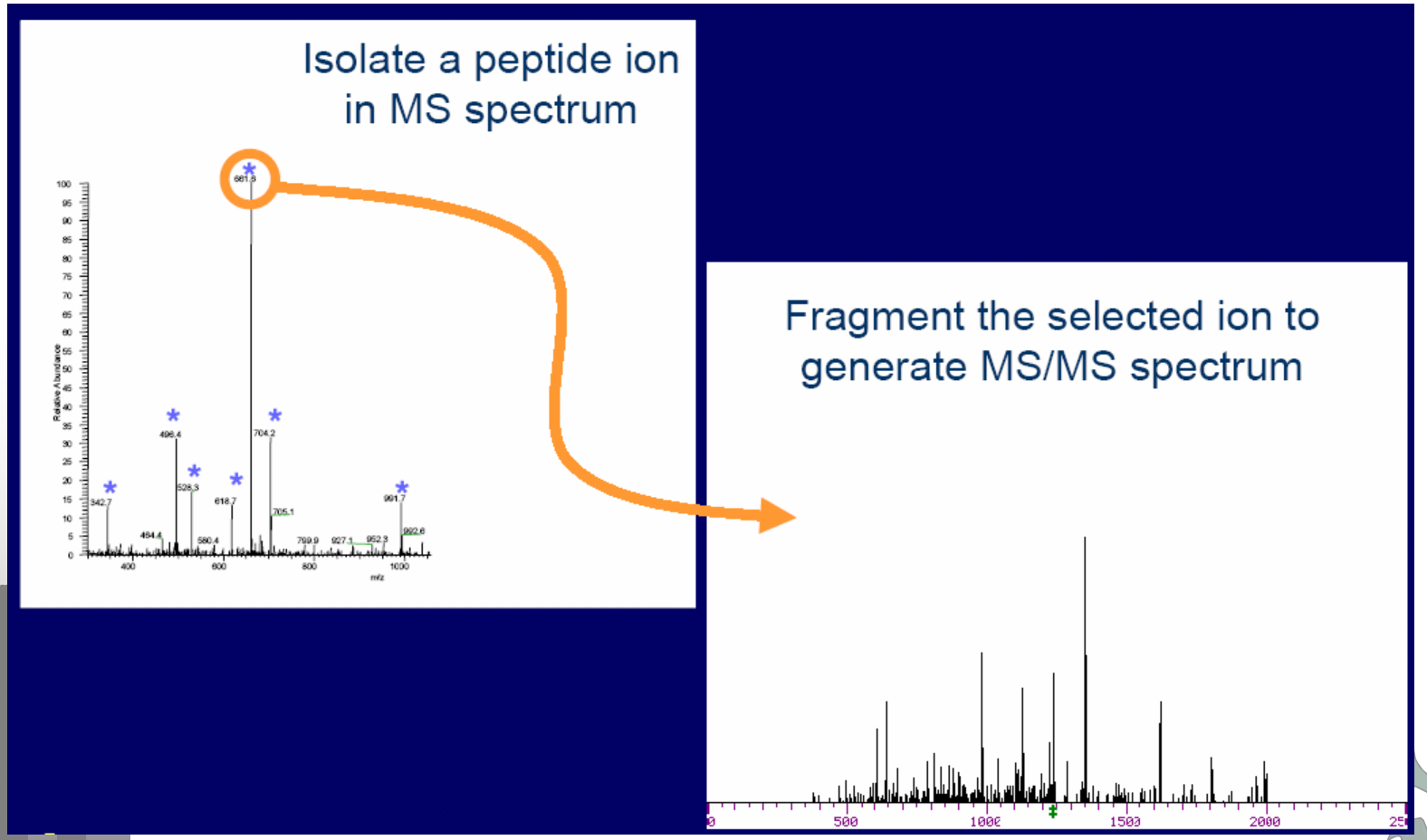
Fenyo (2000) Curr Opinion Biotech 11, 391-395



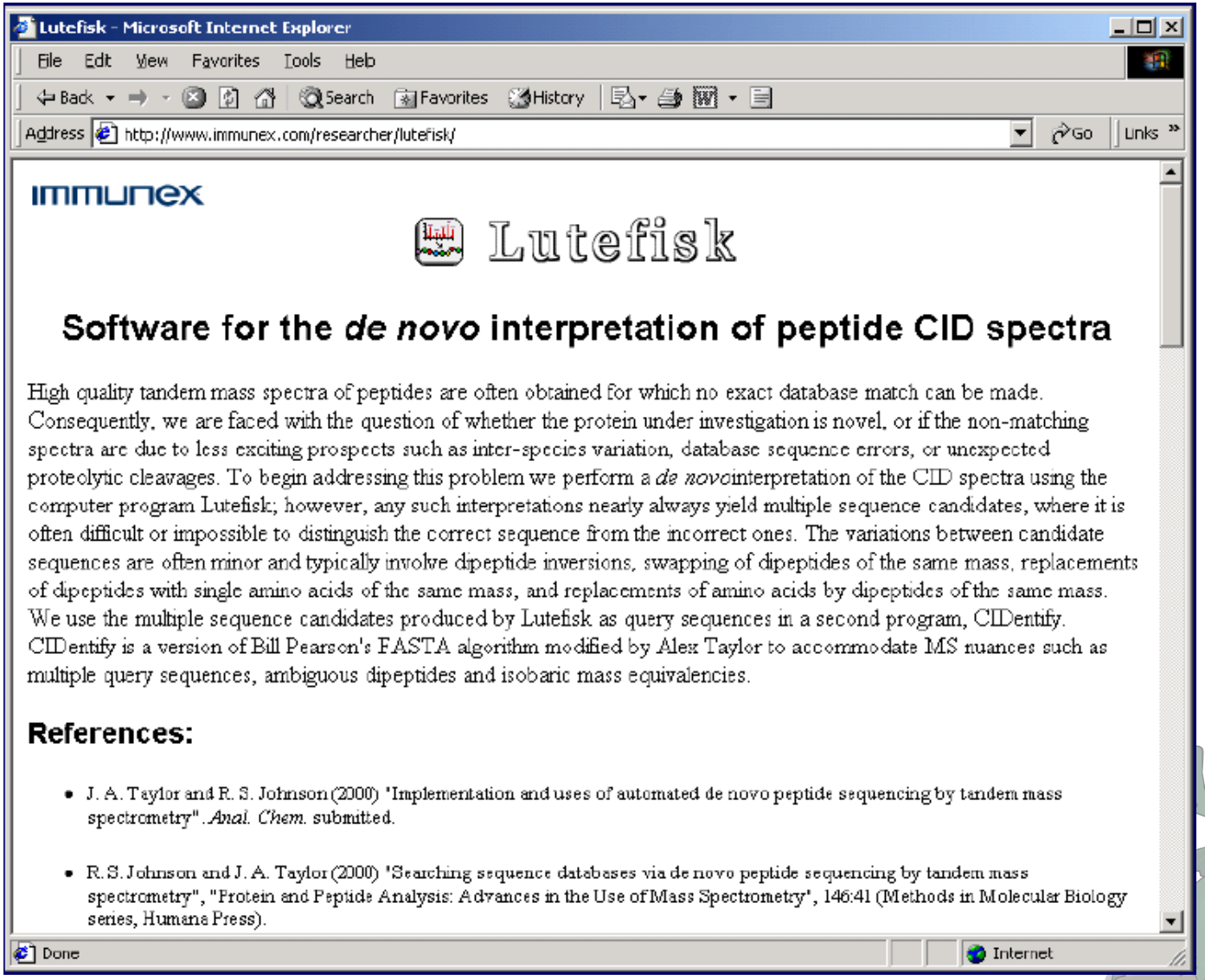
Single tryptic peptide at a mass accuracy of 0.5 dalton

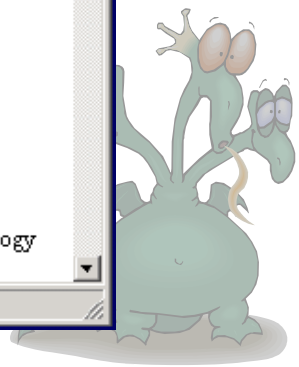
E coli ~ 4000 ORFs
Yeast ~ 6000 ORFs
Human ~ 100,000 ORFs

2nd type: MS/MS data



Another free program

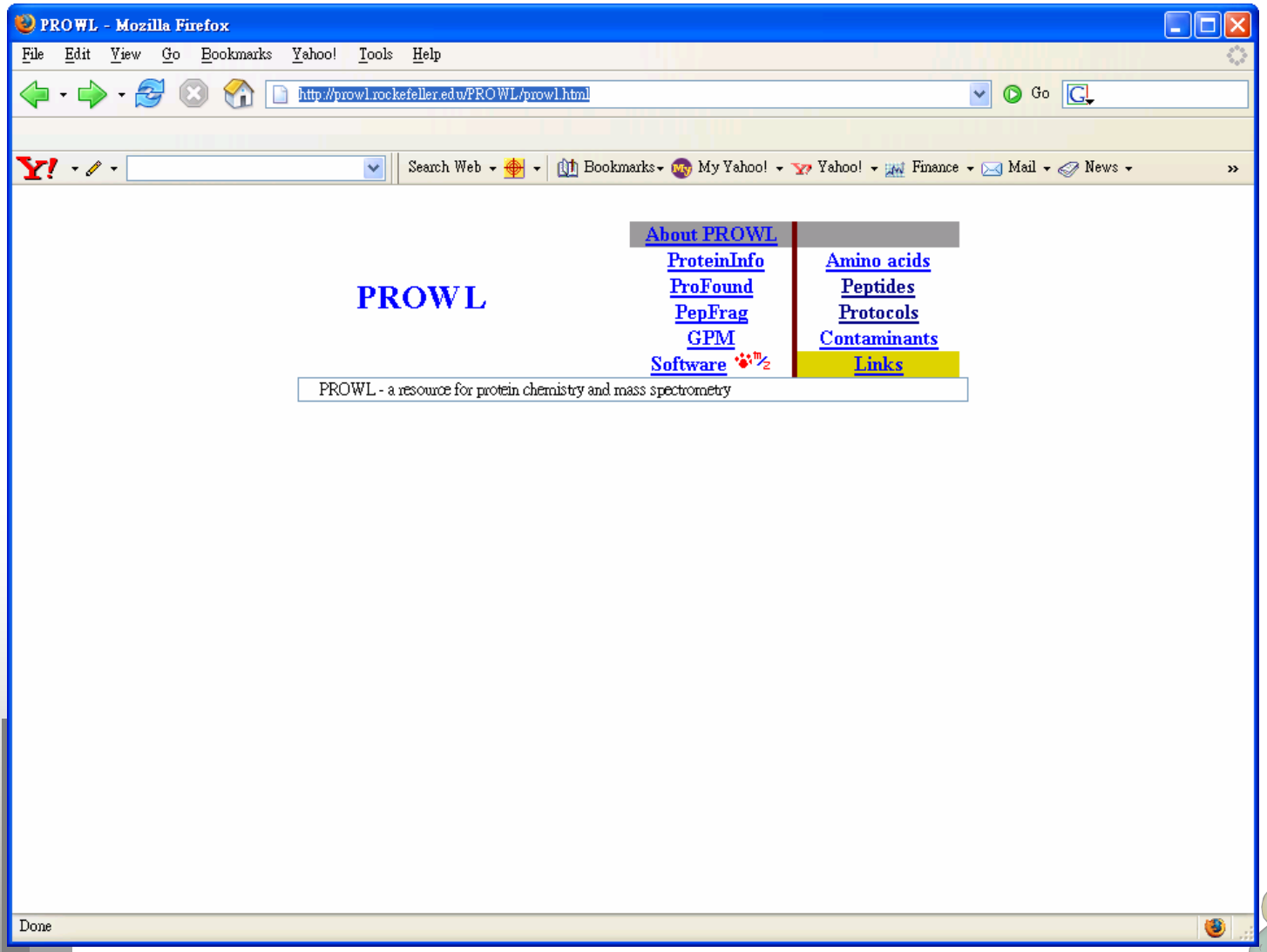


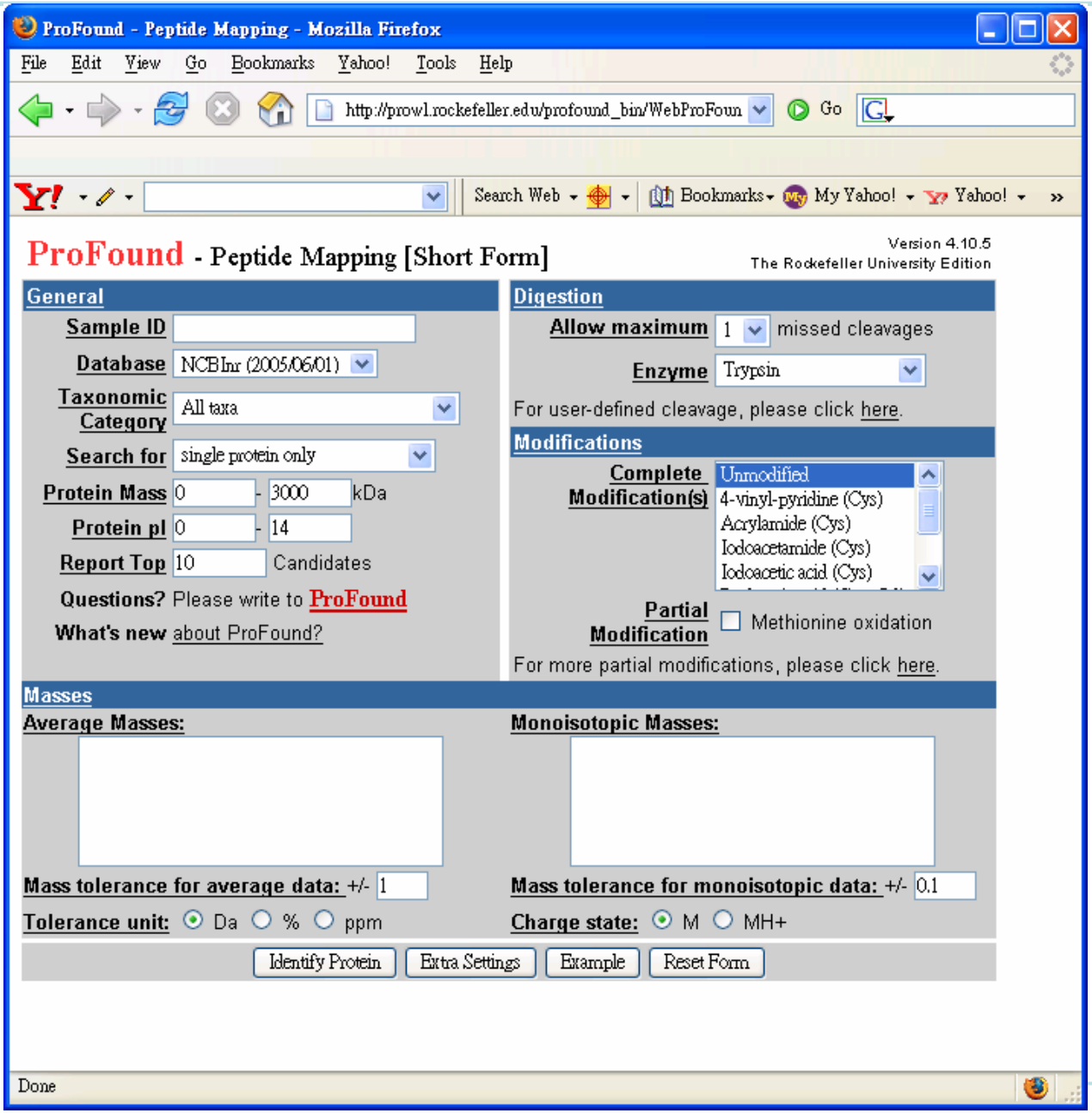


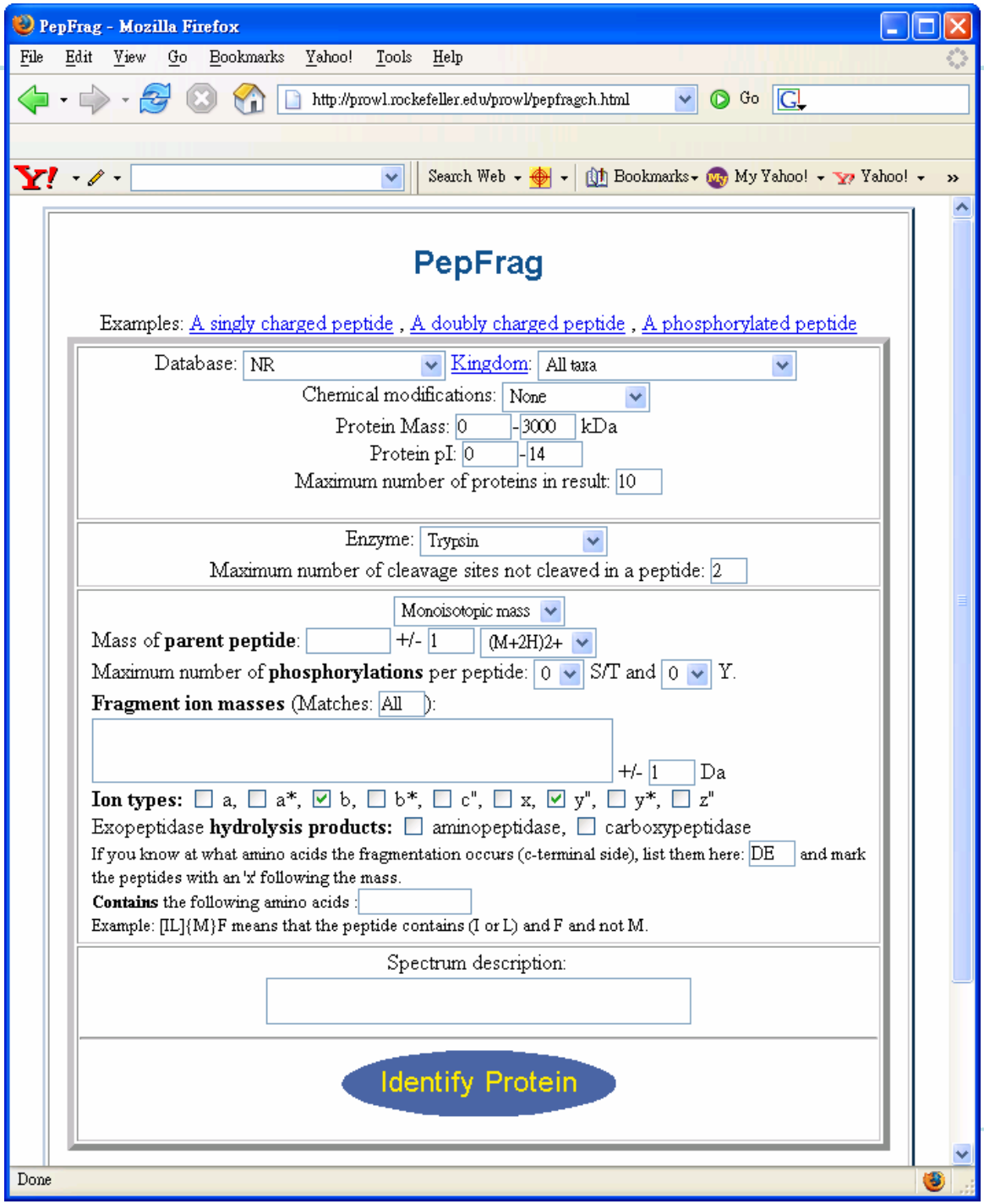
On Web resources:

Prowl at Rockefeller U

http://prowl.rockefeller.edu/PROWL/prowl.html

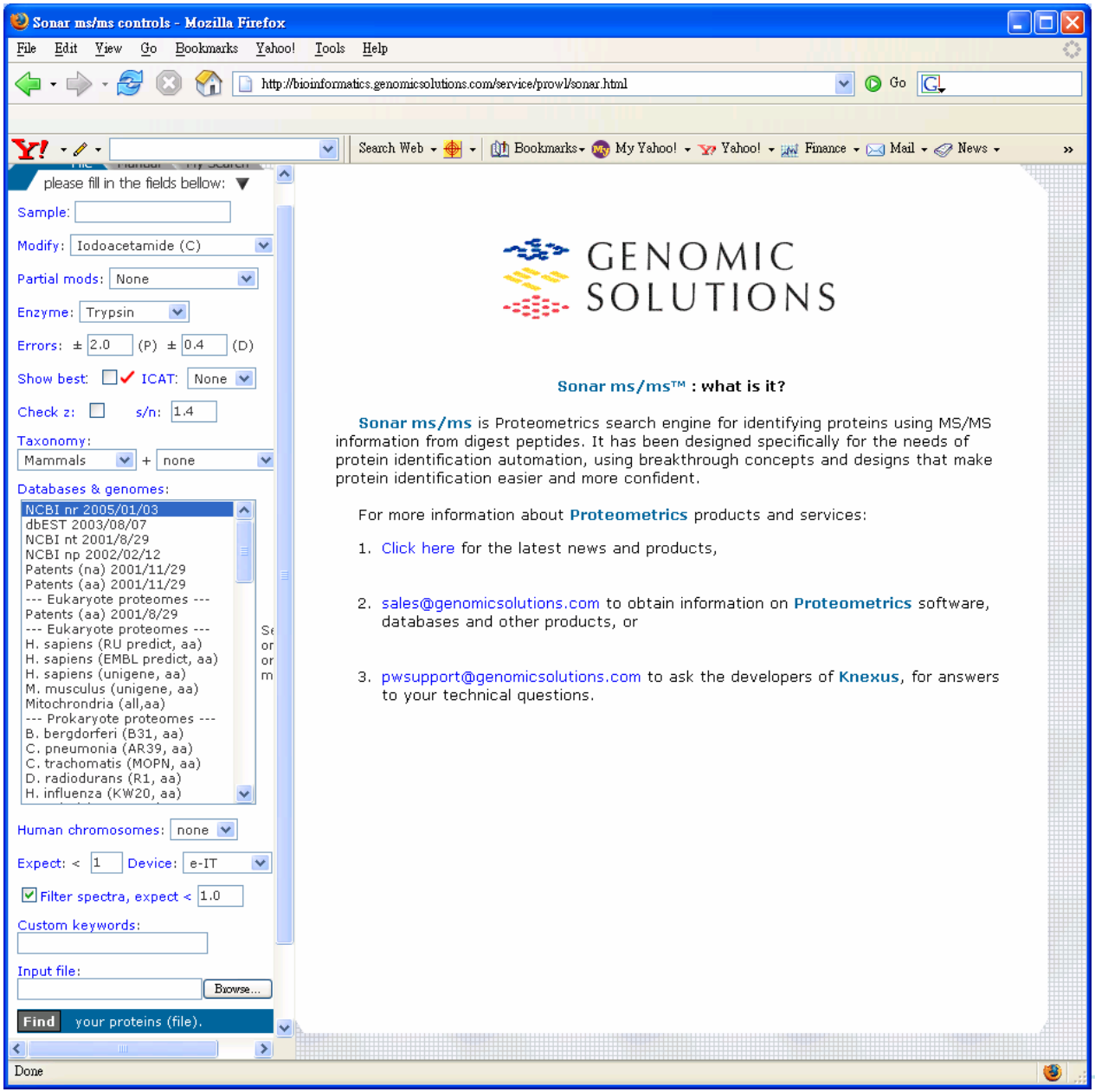

PROWL - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://prowl.rockefeller.edu/PROWL/prowl.html
Go
Search Web
Bookmarks
My Yahoo!
Yahoo!
Finance
Mail
News
PROWL
About PROWL
ProteinInfo
ProFound
PepFrag
GPM
Software
Amino acids
Peptides
Protocols
Contaminants
Links
PROWL - a resource for protein chemistry and mass spectrometry
Done


ProFound - Peptide Mapping - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://prowl.rockefeller.edu/profound_bin/WebProFoun
Go
Search Web
Bookmarks
My Yahoo!
Yahoo!
ProFound - Peptide Mapping [Short Form]
Version 4.10.5
The Rockefeller University Edition
General
Sample ID
Database
NCBInr (2005/06/01)
Taxonomic Category
All taxa
Search for
single protein only
Protein Mass
0
3000
kDa
Protein pI
0
14
Report Top
10
Candidates
Questions? Please write to ProFound
What's new about ProFound?
Digestion
Allow maximum
1
missed cleavages
Enzyme
Trypsin
For user-defined cleavage, please click here.
Modifications
Complete Modification(s)
Unmodified
4-vinyl-pyridine (Cys)
Acrylamide (Cys)
Iodoacetamide (Cys)
Iodoacetic acid (Cys)
Partial Modification
Methionine oxidation
For more partial modifications, please click here.
Masses
Average Masses:
Monoisotopic Masses:
Mass tolerance for average data: +/-
1
Mass tolerance for monoisotopic data: +/-
0.1
Tolerance unit:
Da
%
ppm
Charge state:
M
MH+
Identify Protein
Extra Settings
Example
Reset Form
Done


PepFrag - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://prowl.rockefeller.edu/prowl/pepfragch.html
Go
Search Web
Bookmarks
My Yahoo!
Yahoo!
PepFrag
Examples: A singly charged peptide , A doubly charged peptide , A phosphorylated peptide
Database: NR
Kingdom: All taxa
Chemical modifications: None
Protein Mass: 0 -3000 kDa
Protein pI: 0 -14
Maximum number of proteins in result: 10
Enzyme: Trypsin
Maximum number of cleavage sites not cleaved in a peptide: 2
Monoisotopic mass
Mass of parent peptide: +/- 1 (M+2H)2+
Maximum number of phosphorylations per peptide: 0 S/T and 0 Y.
Fragment ion masses (Matches: All):
+/- 1 Da
Ion types: a, a*, b, b*, c", x, y", y*, z"
Exopeptidase hydrolysis products: aminopeptidase, carboxypeptidase
If you know at what amino acids the fragmentation occurs (c-terminal side), list them here: DE and mark the peptides with an 'x' following the mass.
Contains the following amino acids :
Example: [IL]{M}F means that the peptide contains (I or L) and F and not M.
Spectrum description:
Identify Protein
Done

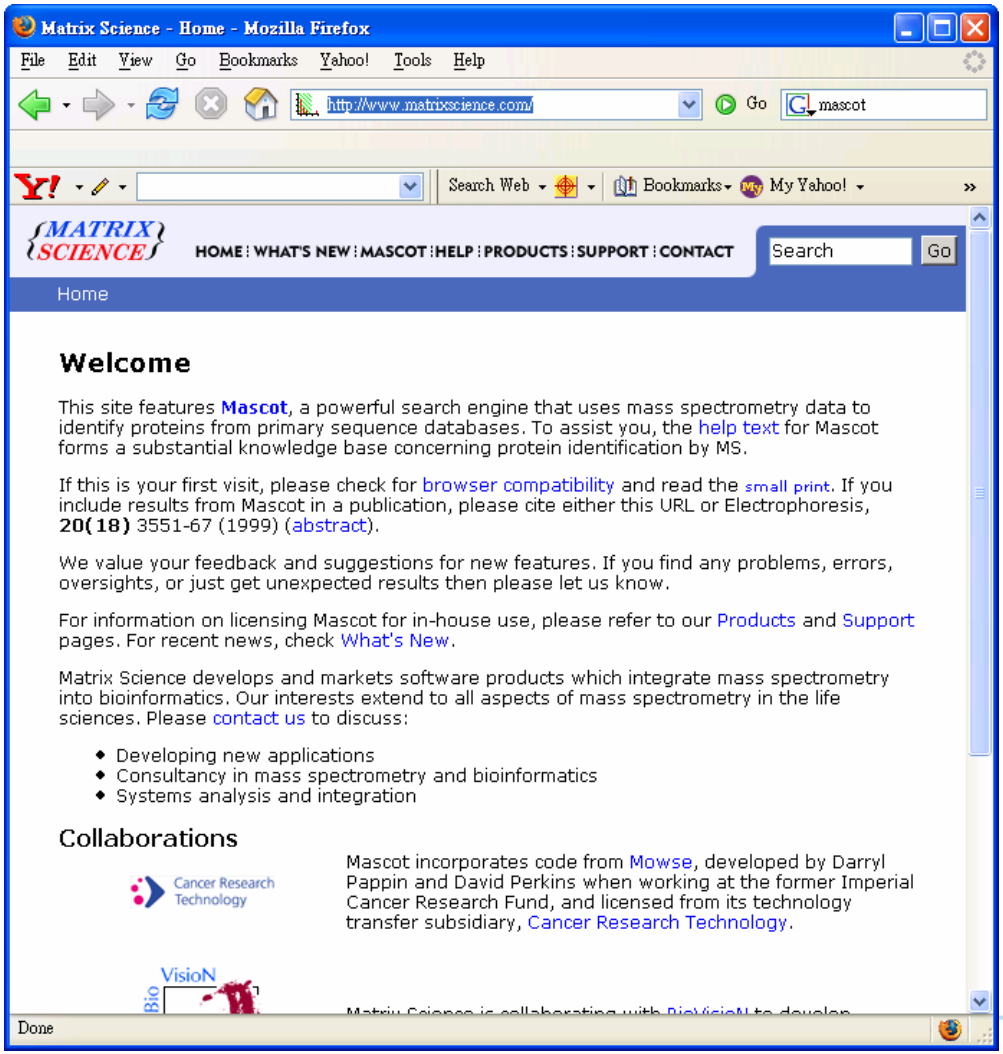
Genomic Solutions

http://www.genomicsolutions.com/showPage.php


Sonar ms/ms controls - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://bioinformatics.genomicsolutions.com/service/prowl/sonar.html
Go
Search Web Bookmarks My Yahoo! Yahoo! Finance Mail News
please fill in the fields bellow:
Sample:
Modify: Iodoacetamide (C)
Partial mods: None
Enzyme: Trypsin
Errors: ± 2.0 (P) ± 0.4 (D)
Show best: ICAT: None
Check z: s/n: 1.4
Taxonomy:
Mammals + none
Databases & genomes:
NCBI nr 2005/01/03
dbEST 2003/08/07
NCBI nt 2001/8/29
NCBI np 2002/02/12
Patents (na) 2001/11/29
Patents (aa) 2001/11/29
--- Eukaryote proteomes ---
Patents (aa) 2001/8/29
--- Eukaryote proteomes ---
H. sapiens (RU predict, aa)
H. sapiens (EMBL predict, aa)
H. sapiens (unigene, aa)
M. musculus (unigene, aa)
Mitochrondria (all,aa)
--- Prokaryote proteomes ---
B. bergdorferi (B31, aa)
C. pneumonia (AR39, aa)
C. trachomatis (MOPN, aa)
D. radiodurans (R1, aa)
H. influenza (KW20, aa)
Human chromosomes: none
Expect: < 1 Device: e-IT
Filter spectra, expect < 1.0
Custom keywords:
Input file:
Browse...
Find your proteins (file).
GENOMIC SOLUTIONS
Sonar ms/ms™ : what is it?
Sonar ms/ms is Proteometrics search engine for identifying proteins using MS/MS information from digest peptides. It has been designed specifically for the needs of protein identification automation, using breakthrough concepts and designs that make protein identification easier and more confident.
For more information about Proteometrics products and services:
1. Click here for the latest news and products,
2. sales@genomicsolutions.com to obtain information on Proteometrics software, databases and other products, or
3. pwsupport@genomicsolutions.com to ask the developers of Knexus, for answers to your technical questions.
Done

Mascot at Matrix-Science

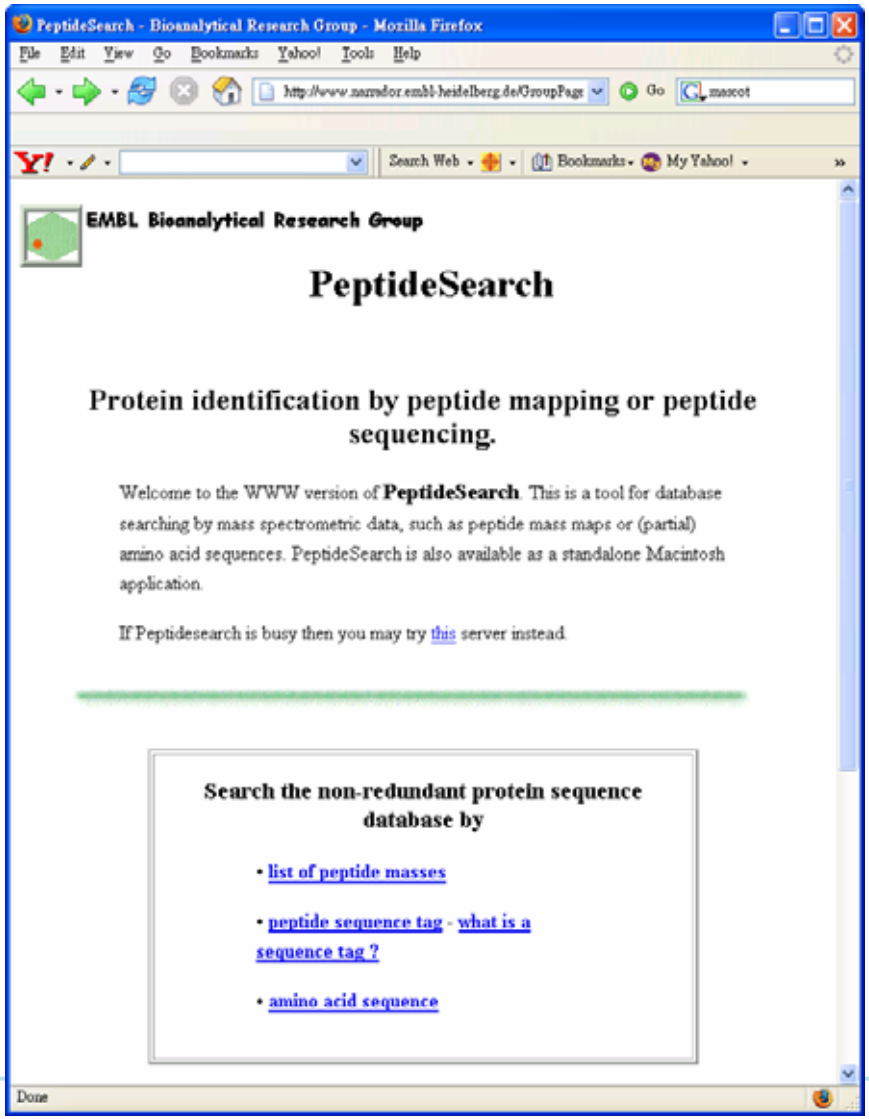
http://www.matrixscience.com/

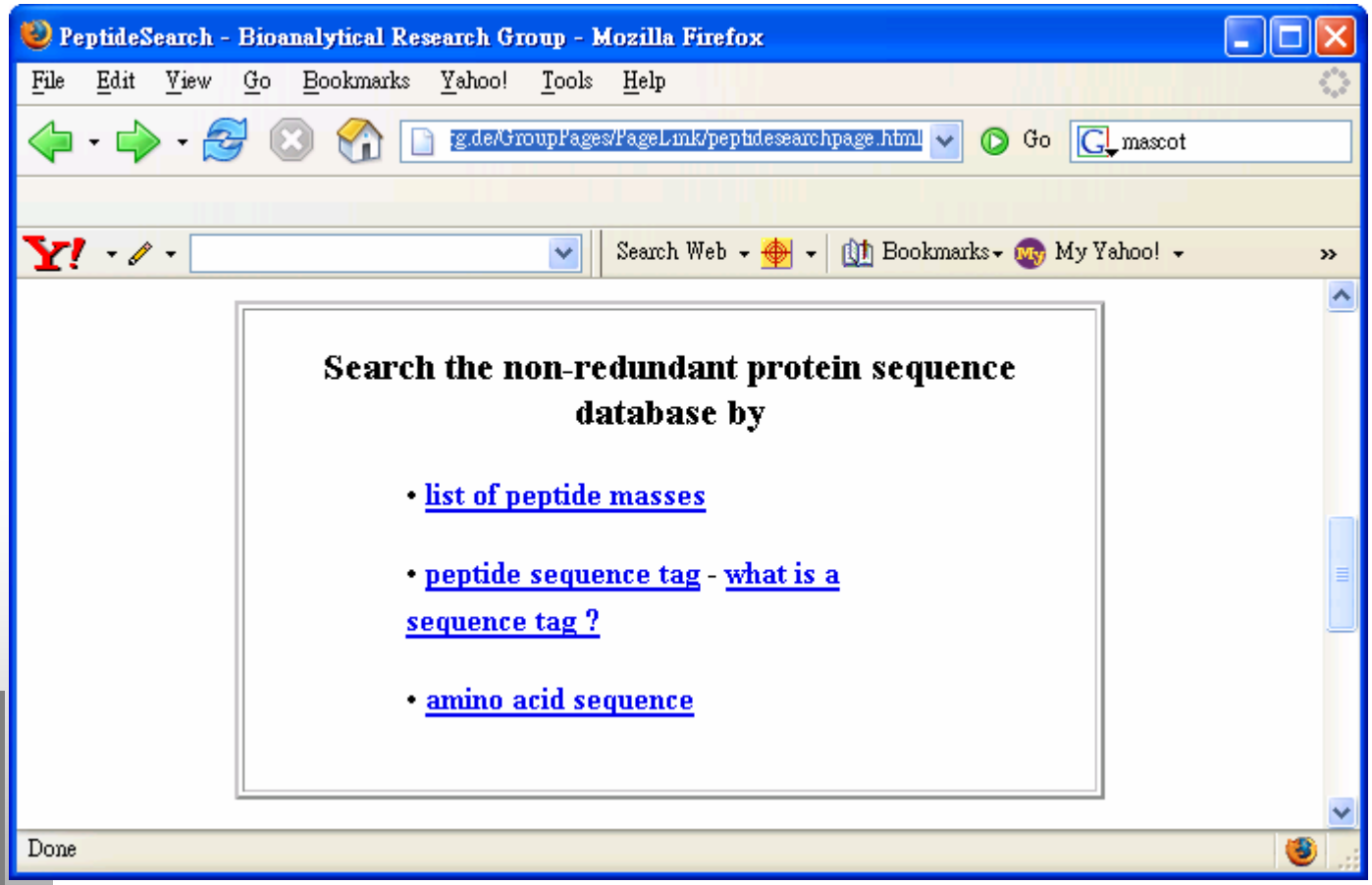



Matrix Science - Mascot - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://www.matrixscience.com/search_form_select.h
Go
mascot
Search Web
Bookmarks
My Yahoo!
MATRIX SCIENCE
HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | CONTACT
Search
Go
Mascot
Mascot Search
Peptide Mass Fingerprint: The experimental data are a list of peptide mass values from an enzymatic digest of a protein.
Example of results report
More information
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. A super-set of a sequence tag query.
Example of results report
More information
MS/MS Ion Search: Identification based on raw MS/MS data from one or more peptides.
Example of results report
More information
Search Form Defaults: Follow this link to save your preferred search form defaults as a browser cookie.
Mascot Help
Mascot Overview
Search parameter reference
Sequence databases
Data file format
Scoring algorithm
Results format
Error tolerant search
FAQ's
User Meeting Presentations
2005
More Help
Help Topic Index
Useful Links
Done

EMBL

http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html



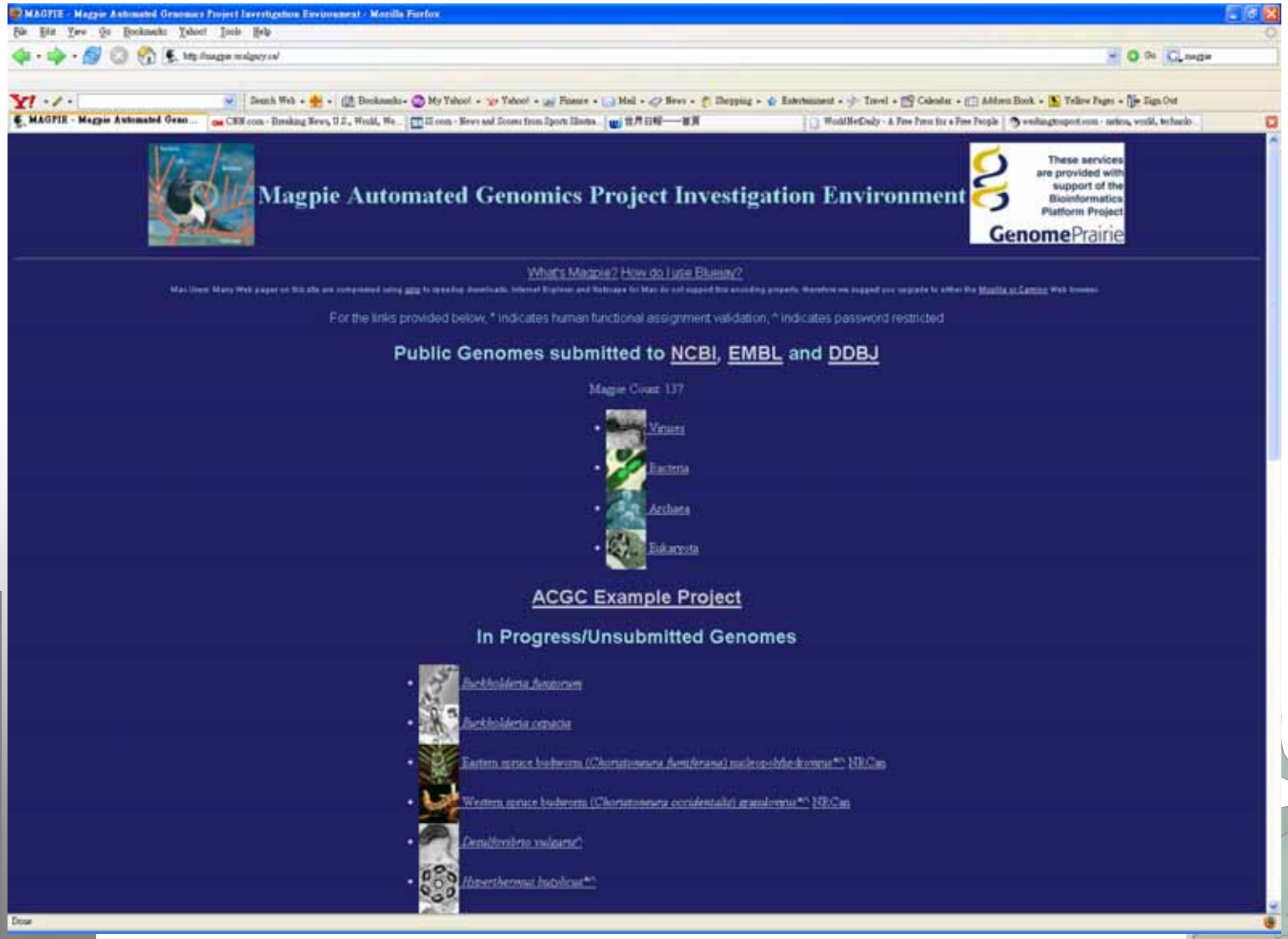

PeptideSearch - Bioanalytical Research Group - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
g.de/GroupPages/PageLink/peptidesearchpage.html
Go
mascot
Search Web
Bookmarks
My Yahoo!
Search the non-redundant protein sequence database by
• list of peptide masses
• peptide sequence tag - what is a sequence tag ?
• amino acid sequence
Done

Blast2p

http://dove.embl-heidelberg.de/Blast2/msblast.html

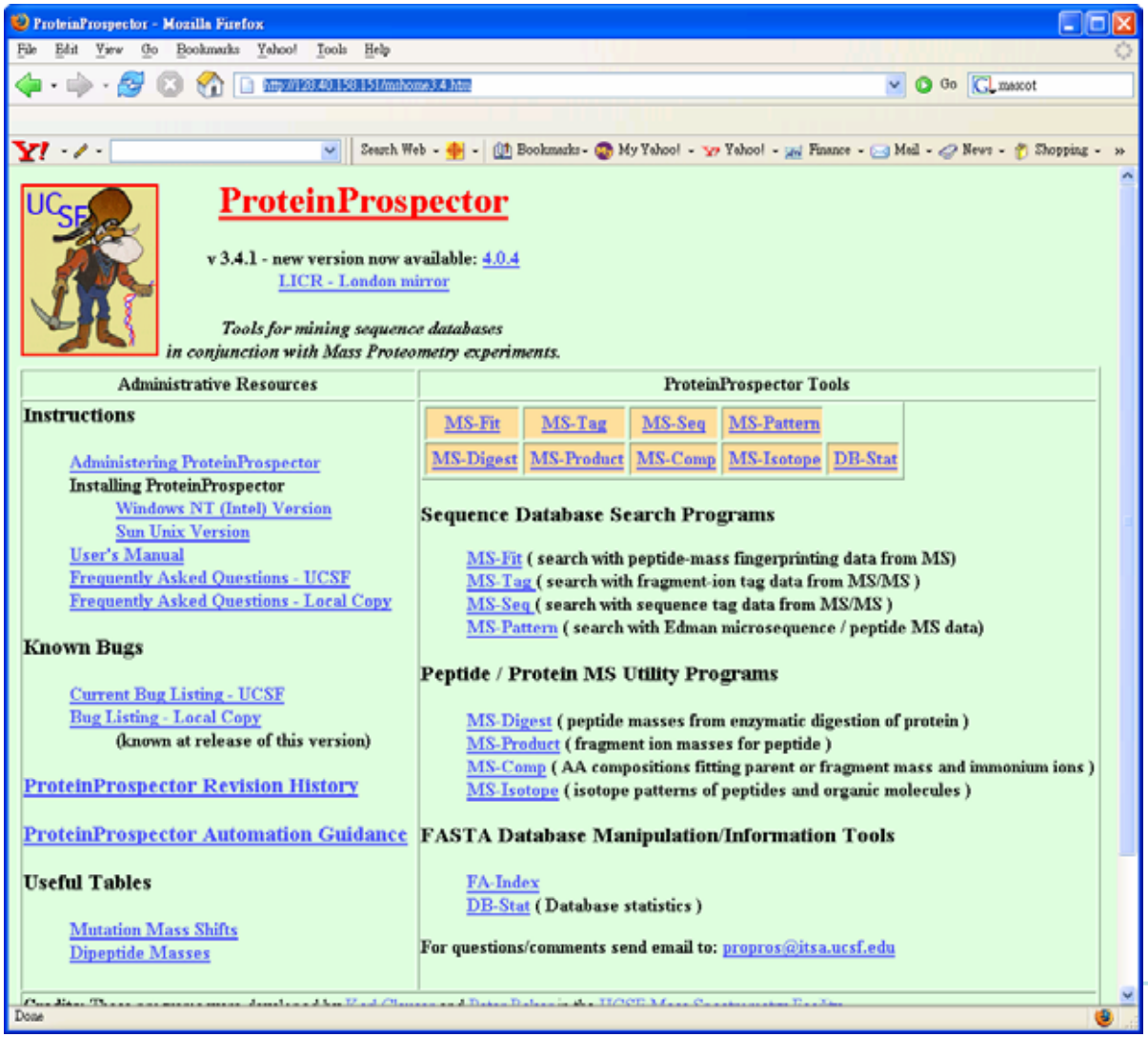
Homology search with partial sequence from ms/ms data

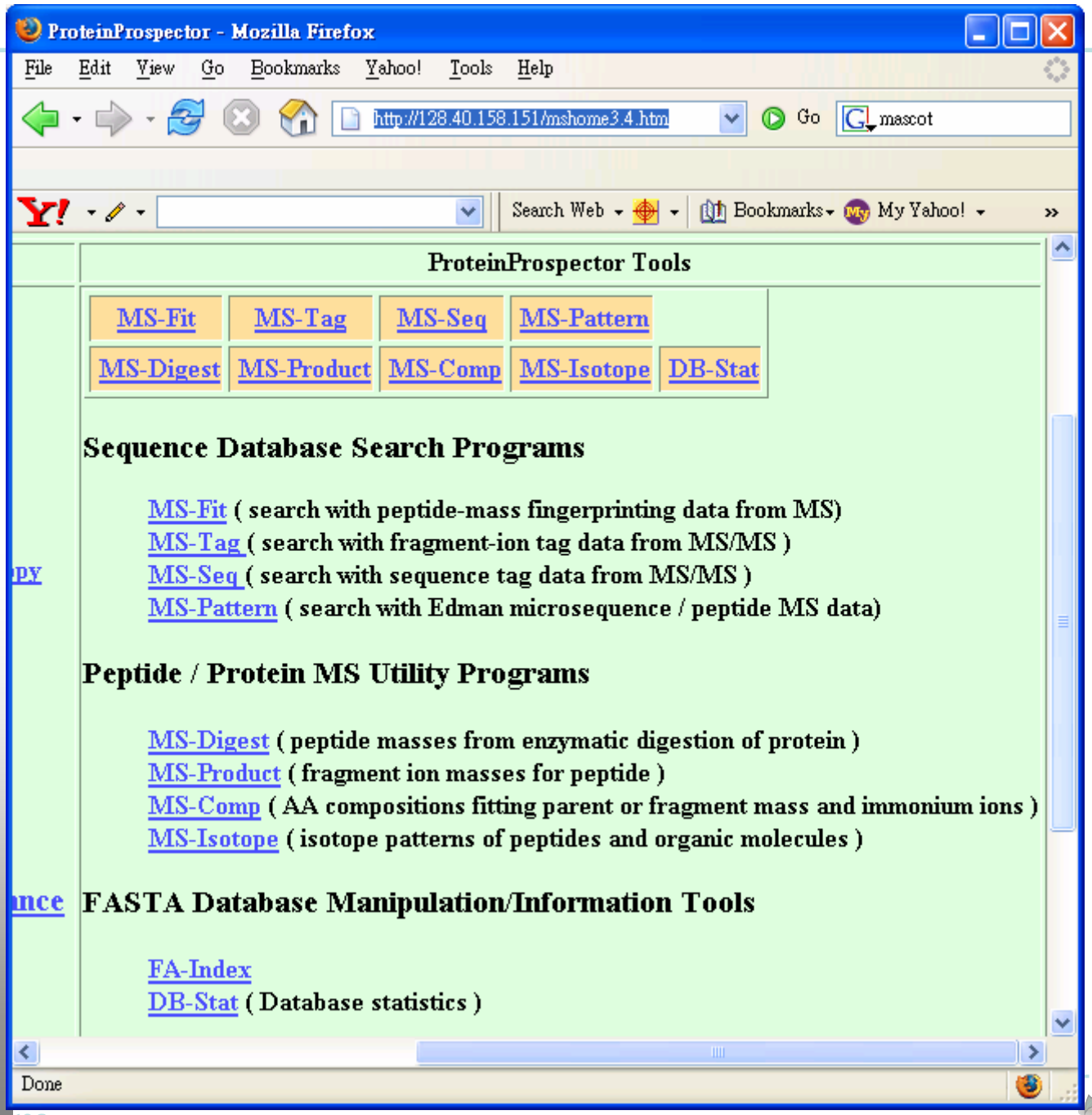
http://magpie.ucalgary.ca/



Prospector at UCSF

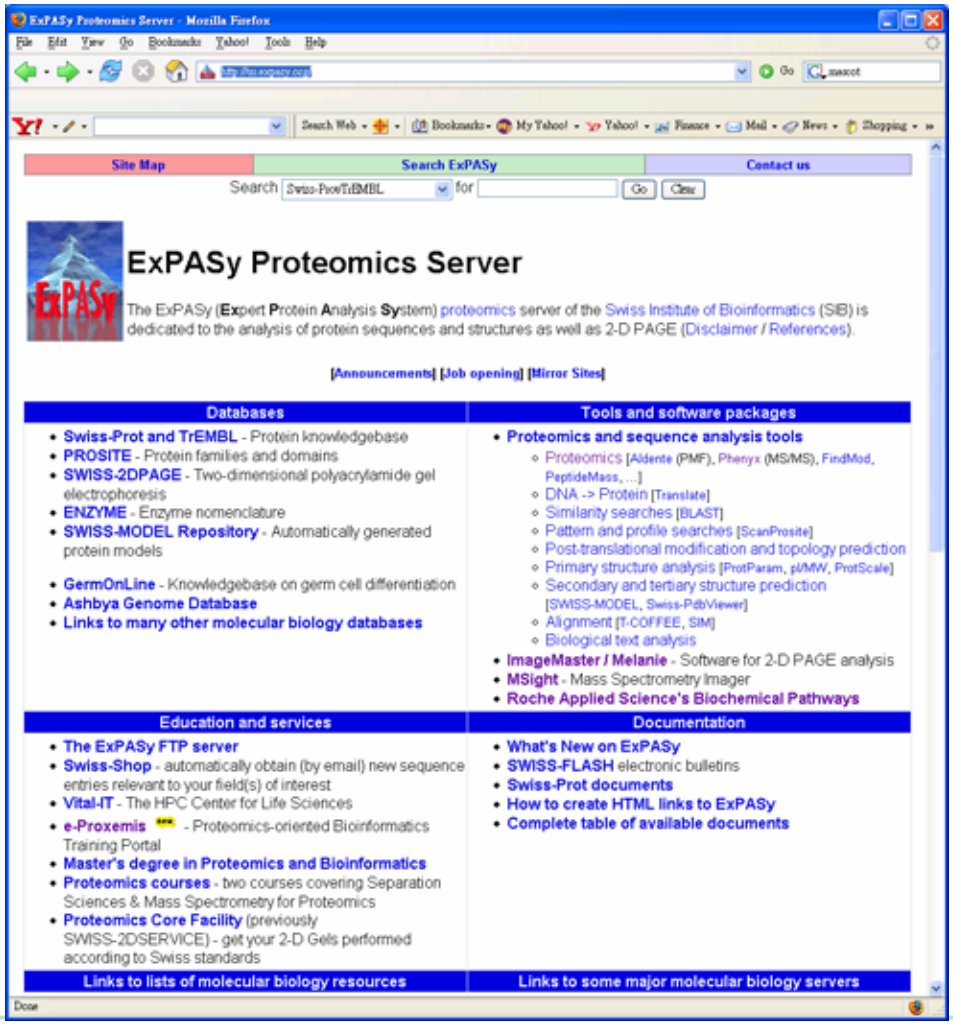
http://128.40.158.151/mshome3.4.htm

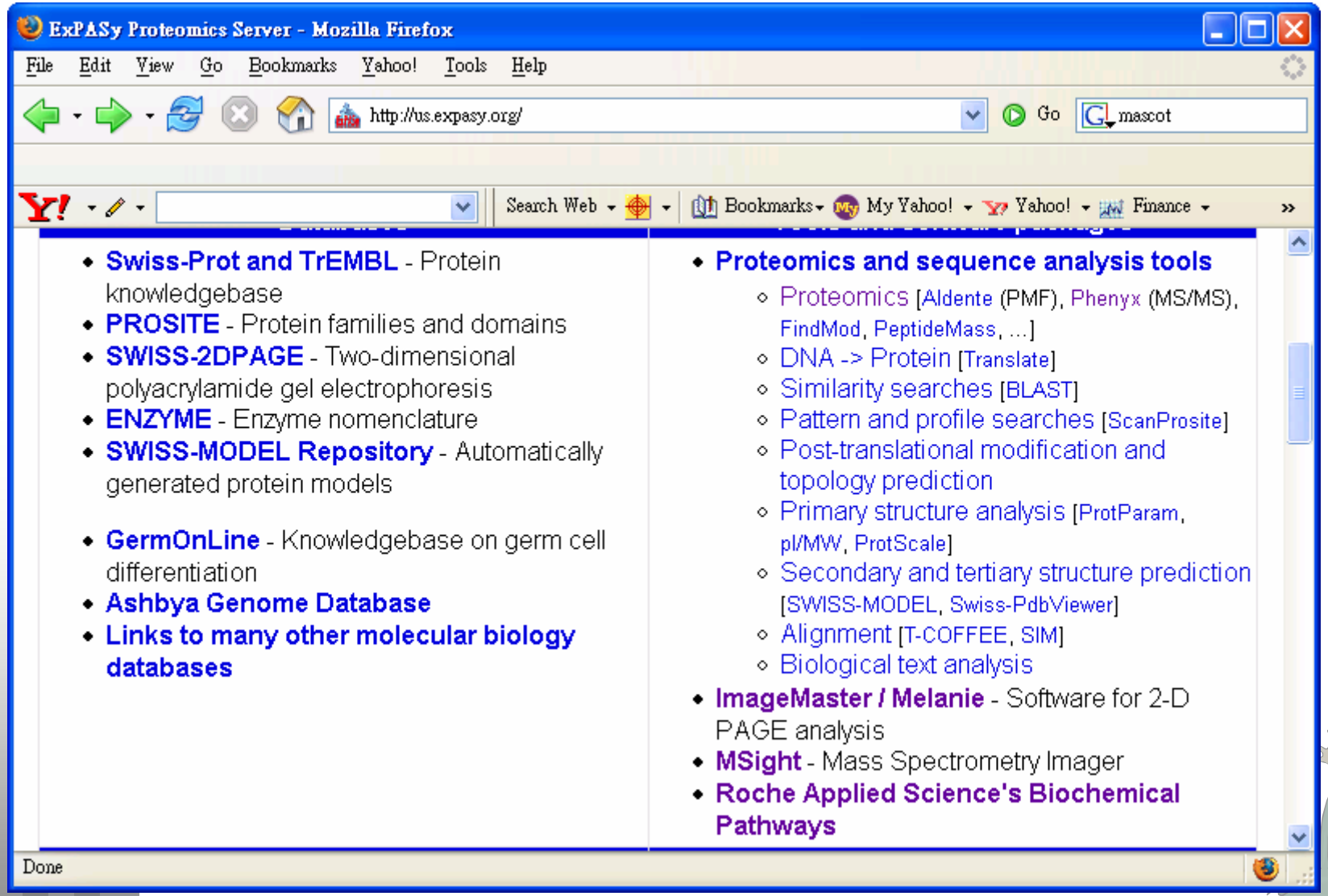



ProteinProspector - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://128.40.158.151/mshome3.4.htm
Go
mascot
Search Web
Bookmarks
My Yahoo!
ProteinProspector Tools
MS-Fit
MS-Tag
MS-Seq
MS-Pattern
MS-Digest
MS-Product
MS-Comp
MS-Isotope
DB-Stat
Sequence Database Search Programs
MS-Fit (search with peptide-mass fingerprinting data from MS)
MS-Tag (search with fragment-ion tag data from MS/MS)
MS-Seq (search with sequence tag data from MS/MS)
MS-Pattern (search with Edman microsequence / peptide MS data)
Peptide / Protein MS Utility Programs
MS-Digest (peptide masses from enzymatic digestion of protein)
MS-Product (fragment ion masses for peptide)
MS-Comp (AA compositions fitting parent or fragment mass and immonium ions)
MS-Isotope (isotope patterns of peptides and organic molecules)
FASTA Database Manipulation/Information Tools
FA-Index
DB-Stat (Database statistics)
Done

EXPASY at SIB

http://us.expasy.org/




ExPASy Proteomics Server - Mozilla Firefox
File Edit View Go Bookmarks Yahoo! Tools Help
http://us.expasy.org/
Go
mascot
Search Web
Bookmarks
My Yahoo!
Yahoo!
Finance
Swiss-Prot and TrEMBL - Protein knowledgebase
PROSITE - Protein families and domains
SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
ENZYME - Enzyme nomenclature
SWISS-MODEL Repository - Automatically generated protein models
GermOnLine - Knowledgebase on germ cell differentiation
Ashbya Genome Database
Links to many other molecular biology databases
Proteomics and sequence analysis tools
Proteomics [Aldente (PMF), Phenyx (MS/MS), FindMod, PeptideMass, ...]
DNA -> Protein [Translate]
Similarity searches [BLAST]
Pattern and profile searches [ScanProsite]
Post-translational modification and topology prediction
Primary structure analysis [ProtParam, pI/MW, ProtScale]
Secondary and tertiary structure prediction [SWISS-MODEL, Swiss-PdbViewer]
Alignment [T-COFFEE, SIM]
Biological text analysis
ImageMaster / Melanie - Software for 2-D PAGE analysis
MSight - Mass Spectrometry Imager
Roche Applied Science's Biochemical Pathways
Done

Guidelines for “protein identification”

Taylor, GK & Goodlett, DR (2005)

RCM 19, 3420

RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2005; 19: 3420
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rcm.2225

RCM

Letter to the Editor

To the Editor-in-Chief
Sir,

Rules governing protein identification by mass spectrometry

On behalf of the Editors we write with regard to our discussion of the need for rules governing protein identification by mass spectrometry published in *Rapid Communications in Mass Spectrometry* (RCM). While protein identification by mass spectrometry is well established, rules governing exactly what parameters constitute identification are not. We suggest here rules that we hope are not overly burdensome on authors, but provide readers with assurance of quality.

The following information should be included in manuscripts that describe 'identification' of proteins by mass spectrometry:

1. The name of programs used to convert raw MS and MS/MS data into 'database searchable' files.
2. The name of the software used to query a sequence database using MS data, e.g. SEQUEST,[1] Mascot,[2] XTandem,[3] or Peaks.[4]
3. The name of the database searched, including any specific time stamp date if it is public and where it may be found, should be included as a reference. For private (or contrived) databases, a full description of the contents including the number of proteins in the database and the average sequence length should be included.
4. A description of the type of scoring and cutoff criteria used to decide that a set of data (MS or MS/MS) indicates the presence of a protein in the sample. For example, if SEQUEST was used for the database search, then state the XCORR and ΔCORR cutoff values.
5. A measure of the false positive rate. The simplest method to calculate this for MS/MS data is to search the data against a reverse sequence database search.[5] Proteins identified with an estimated false positive rate greater than 10% are considered dubious identifications (see item 11 below).
6. For each protein identified, a list of the peptides matched with their scores and an accounting of protein sequence coverage. Very long lists may need to be supplied as supplementary data, but should be included at the time of submission for review.
7. Proteins identified by a single peptide MS/MS spectrum match are considered dubious identifications (see item 11) and are discouraged.
8. Should identification be based on peptide mass fingerprinting (PMF), mass accuracies of peptides used for the identification should be stated. Proteins identified by PMF using *m/z* peaks with signal/noise less than 1.5 are considered dubious (see item 11).
9. For identifications from either MS/MS or PMF, authors are strongly encouraged to use a software package that assigns an objective statistical basis to the identification, e.g. Protein Prophet.[6]
10. Where peptides used for identification match to homologous proteins, authors must indicate why a particular species was chosen over all possible homologues.
11. Dubious protein/peptide identifications (see items 5, 7 and 8) must be verified by complementary means. This should include the manually annotated tandem mass spectrum as supplementary data, and any corroborating non-MS data such as reaction with protein-specific antibodies in a Western blot or ELISA.

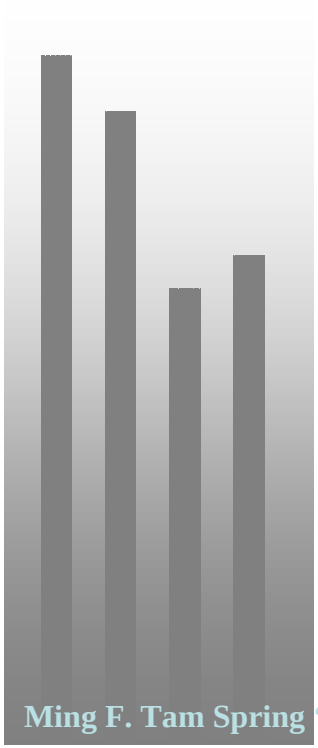
Finally, we point out that it could be argued that data used to justify any publicly available manuscript must be publicly available also. While RCM has no authority beyond the manuscript in its published form to make data public, authors are reminded that many granting agencies are increasingly interested in public availability of data generated with public funds. Thus we remind authors that they may receive requests for raw files that support specific identifications, and we encourage dissemination of this information, as it should increase the quality of their own work as well as aid the community at large.

*Gregory K. Taylor and David R. Goodlett**
University of Washington, Department of Medicinal Chemistry, 1959 NE Pacific Street, Health Sciences Bldg. Box 357610, Seattle, WA 98199, USA

**Correspondence to: D. R. Goodlett, University of Washington, Department of Medicinal Chemistry, 1959 NE Pacific Street, Health Sciences Bldg. Box 357610, Seattle, WA 98199, USA.*
E-mail: goodlett@u.washington.edu

REFERENCES

1. Eng J, McCormack AL, Yates JR. *J. Am. Soc. Mass Spectrom.* 1994; 5: 976.
2. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. *Electrophoresis* 1999; 20: 3551.
3. Fenyo D, Beavis RC. *Anal. Chem.* 2003; 75: 768.
4. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. *Rapid Commun. Mass Spectrom.* 2003; 17: 2337.
5. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. *J. Proteome Res.* 2003; 2: 43.
6. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. *Anal. Chem.* 2002; 74: 5383.



Ming F. Tam Spring '06

lished in *Rapid Communications in Mass Spectrometry* (RCM). While protein identification by mass spectrometry is well established, rules governing exactly what parameters constitute identification are not. We suggest here rules that we hope are not overly burdensome on authors, but provide readers with assurance of quality.

The following information should be included in manuscripts that describe 'identification' of proteins by mass spectrometry:

1. The name of programs used to convert raw MS and MS/MS data into 'database searchable' files.
2. The name of the software used to query a sequence database using MS data, e.g. SEQUEST,[1] Mascot,[2] X!Tandem,[3] or Peaks.[4]
3. The name of the database searched, including any specific time stamp date if it is public and where it may be found, should be included as a reference. For private (or contrived) databases, a full description of the contents including the number of proteins in the database and the average sequence length should be included.

4. A description of the type of scoring and cutoff criteria used to decide that a set of data (MS or MS/MS) indicates the presence of a protein in the sample. For example, if SEQUEST was used for the database search, then state the XCORR and ΔCORR cutoff values.
5. A measure of the false positive rate. The simplest method to calculate this for MS/MS data is to search the data against a reverse sequence database search.[5] Proteins identified with an estimated false positive rate greater than 10% are considered dubious identifications (see item 11 below).

J Proteome Res (2003) 2, 43-50.

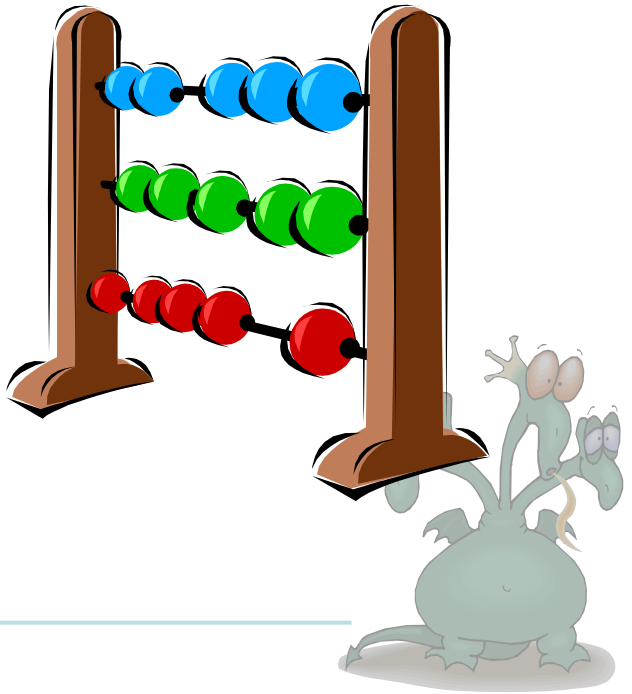
6. For each protein identified, a list of the peptides matched with their scores and an accounting of protein sequence coverage. Very long lists may need to be supplied as supplementary data, but should be included at the time of submission for review.
7. Proteins identified by a single peptide MS/MS spectrum match are considered dubious identifications (see item 11) and are discouraged.

8. Should identification be based on peptide mass fingerprinting (PMF), mass accuracies of peptides used for the identification should be stated. Proteins identified by PMF using *m/z* peaks with signal/noise less than 1.5 are considered dubious (see item 11).
9. For identifications from either MS/MS or PMF, authors are strongly encouraged to use a software package that assigns an objective statistical basis to the identification, e.g. Protein Prophet.[6]

http://www.systemsbiology.org/research/software.html
http://proteinprophet.sourceforge.net/ .

10. Where peptides used for identification match to homologous proteins, authors must indicate why a particular species was chosen over all possible homologues.
11. Dubious protein/peptide identifications (see items 5, 7 and 8) must be verified by complementary manually annotated tandem mass spectrum as supplementary data, and any corroborating non-MS data such as reaction with protein-specific antibodies in a Western blot or ELISA.

Mass Spectrometers are not Quantitative Instruments

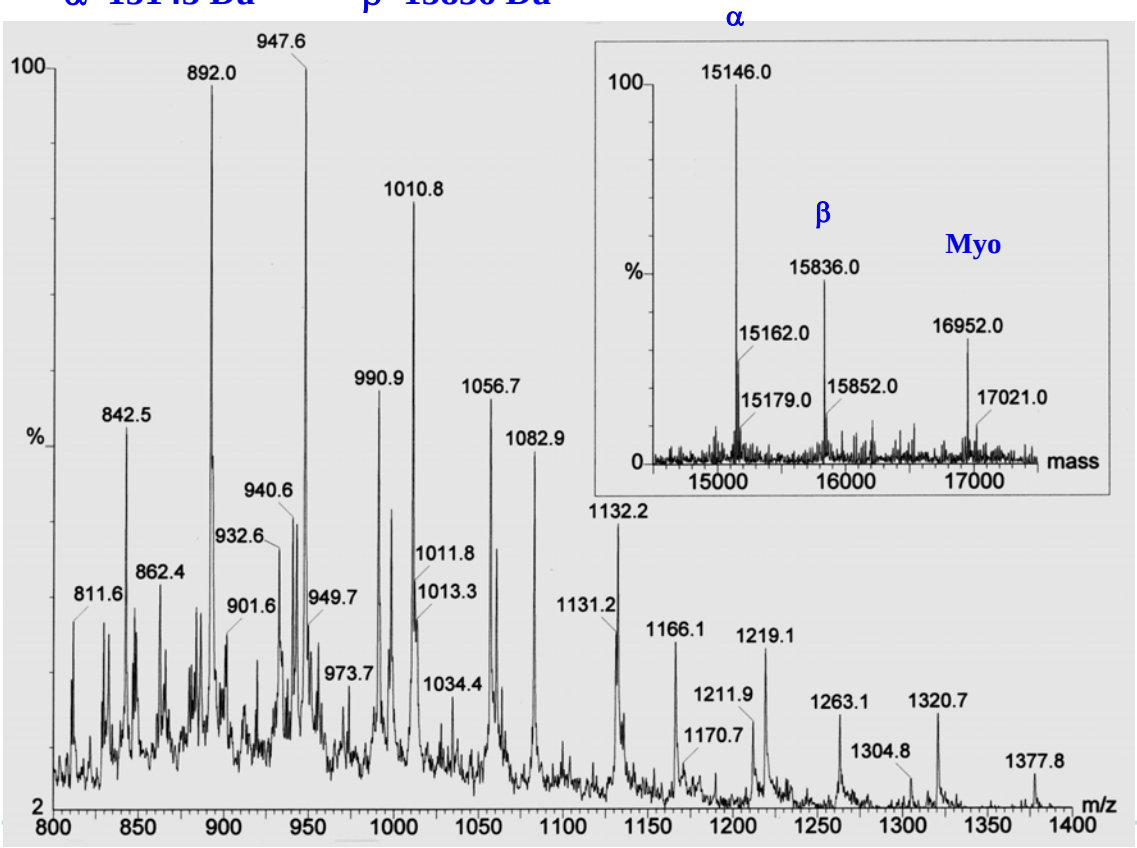


Electrospray source/quadrupole analyzer
Hemoglobin is a tetramer $\alpha_2\beta_2$

[Myo]: [α]: [β] 1: 2.5 : 2.5

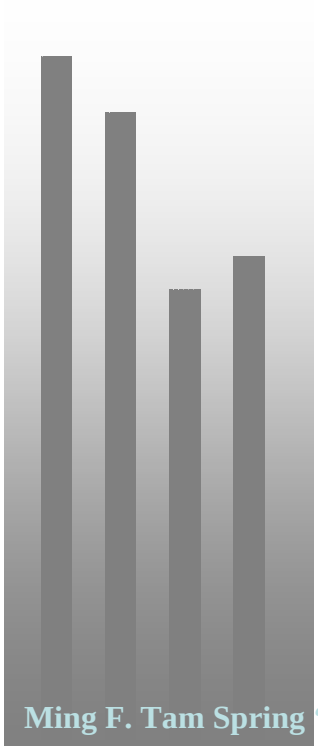
Recombinant Human Hemoglobin αH20R/βE6V, E22Q

α=15145 Da β=15836 Da

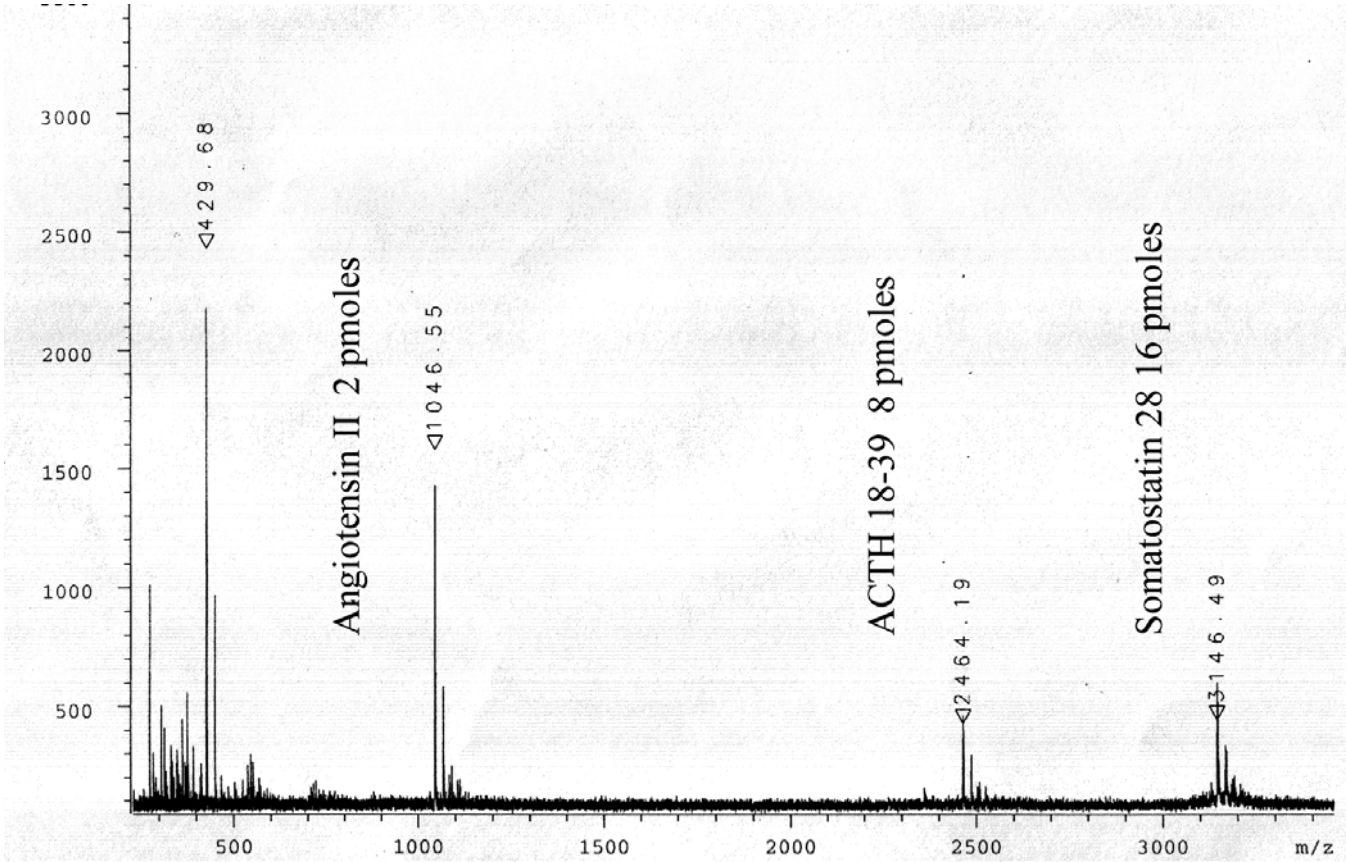


Ionization efficiency and charges on polypeptide

	Lys	Arg	His	Σ
α	11	3	10	24
β	11	3	9	23
Myo	19	2	11	33



MALDI-TOF instrument



Angiotensin II

D-R-V-Y-I-H-P-F

ACTH 18-39

RPVKVYPNGAEDESAEAFPLEF

Somatostatin 28

SANSNPAMAPRERKAGCKNFFWKTFTSC

Stone age technique

- 2 DE and image analysis of gels
- Staining method (Coomassie, silver and SYPRO dyes)
  - Sensitivity
  - Linearity of the staining range,
  - Gel to gel variation

General approach:

Put a tag on (modify) the proteins. Compare the tagged and the untagged samples.

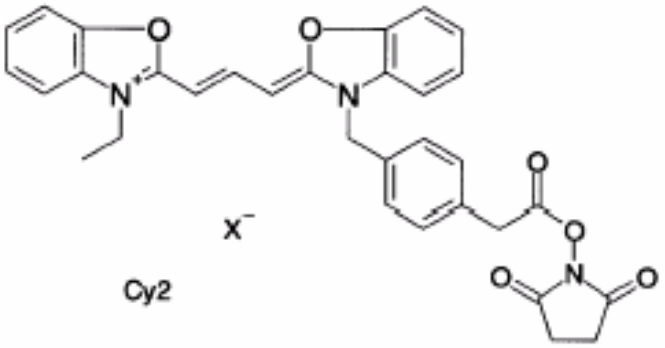
Not too many choices—tags have to be on reactive sidechains:

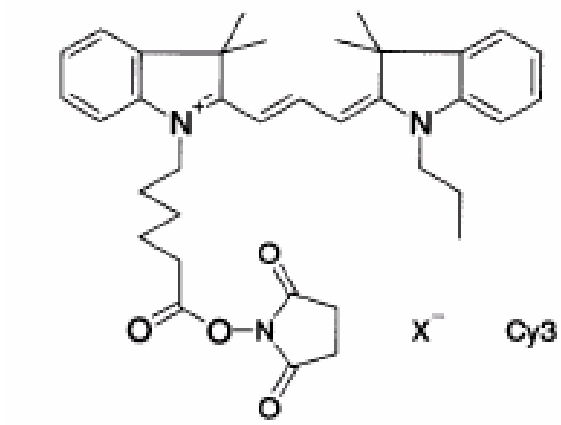
Lys	His	Arg
Cys	Trp	Gly, Ala, Val, Lxx
Met	Tyr	Asn, Gln, Phe, Pro
Ser/Thr(?)	Glu/Asp	

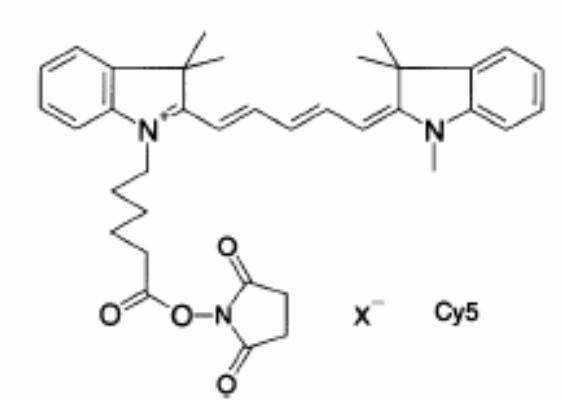
In vitro labeling technique

Difference Gel Electrophoresis (DIGE) system

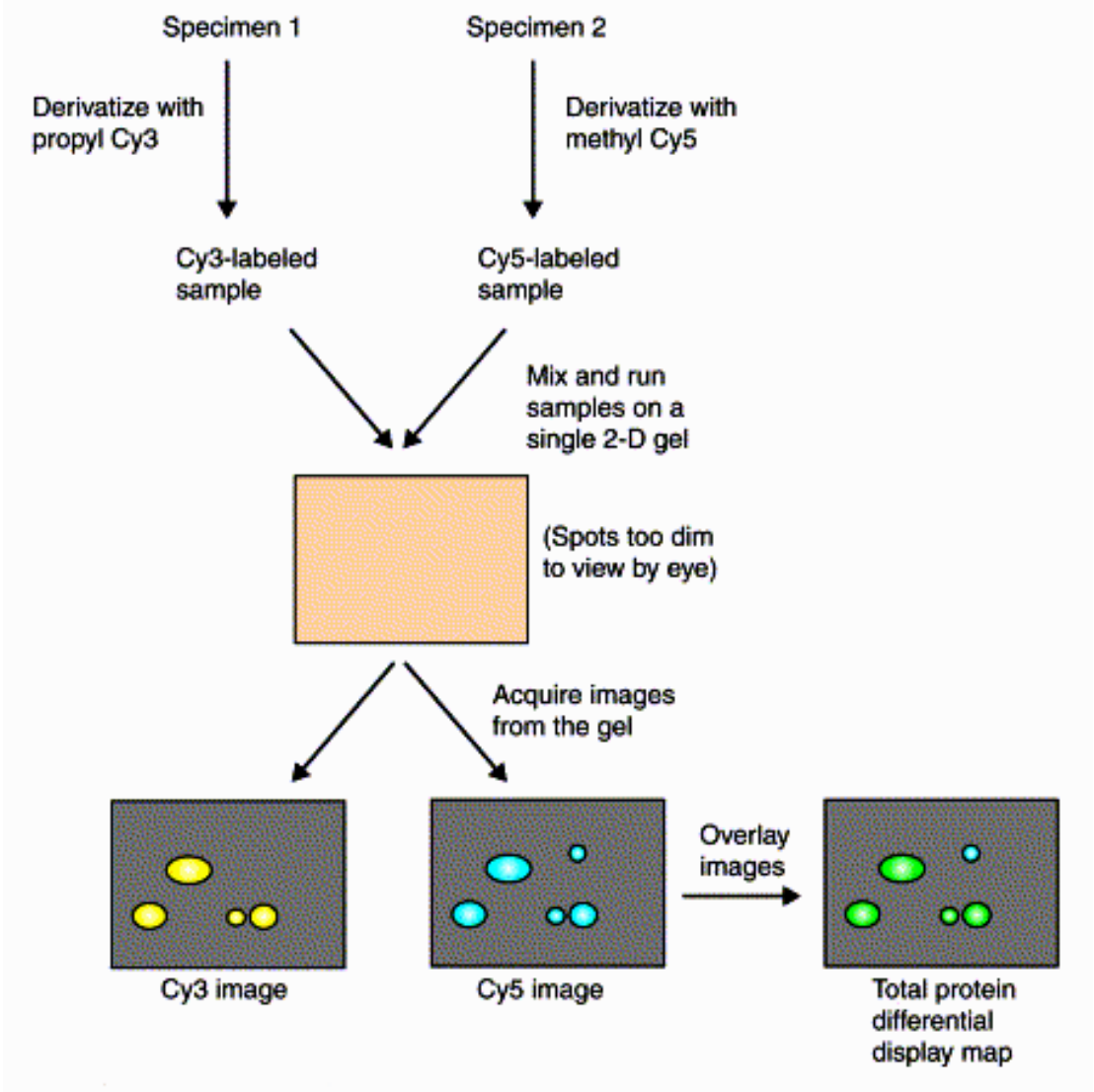
Ronge et al. *Proteomics* 2001, 6:117-128







Modification of lysine side chains

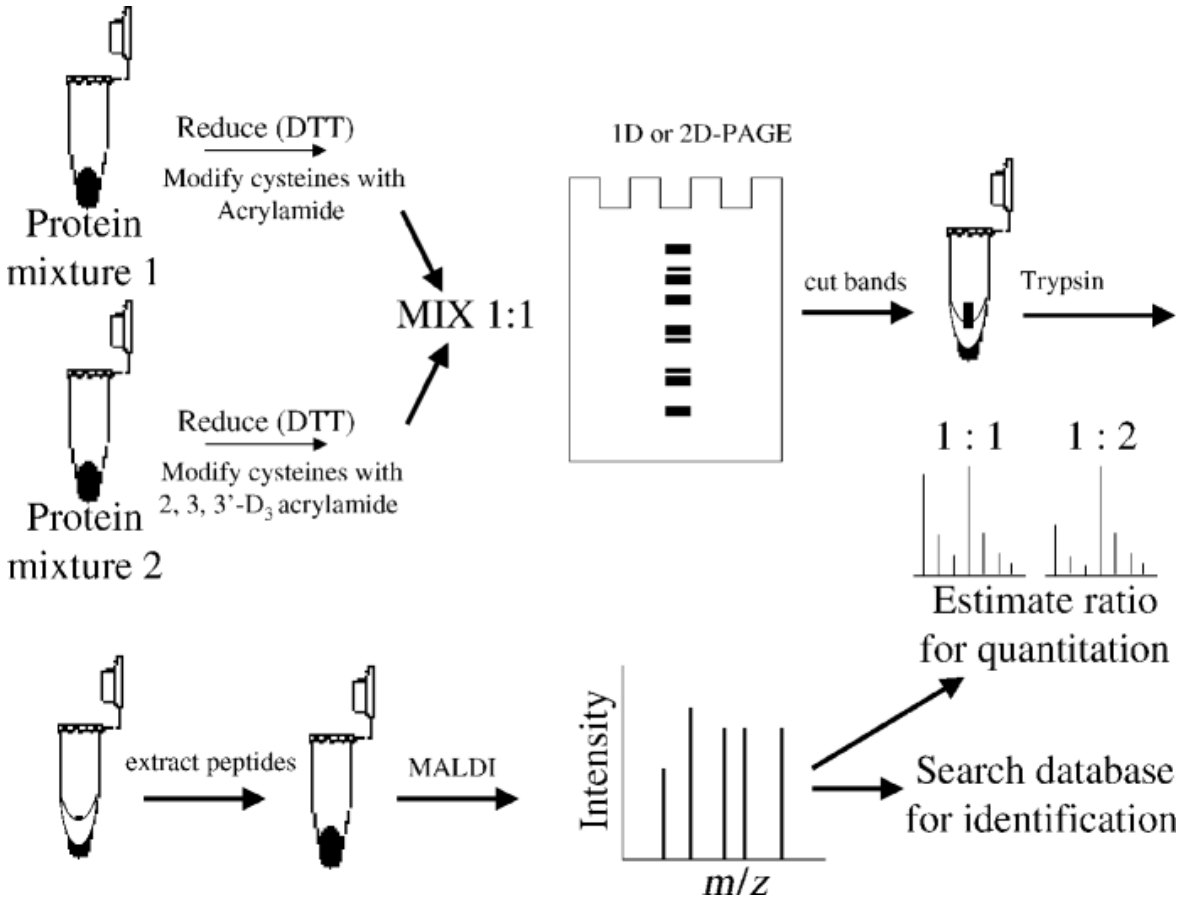

Specimen 1
Specimen 2
Derivatize with propyl Cy3
Derivatize with methyl Cy5
Cy3-labeled sample
Cy5-labeled sample
Mix and run samples on a single 2-D gel
(Spots too dim to view by eye)
Acquire images from the gel
Cy3 image
Cy5 image
Overlay images
Total protein differential display map

Problems:

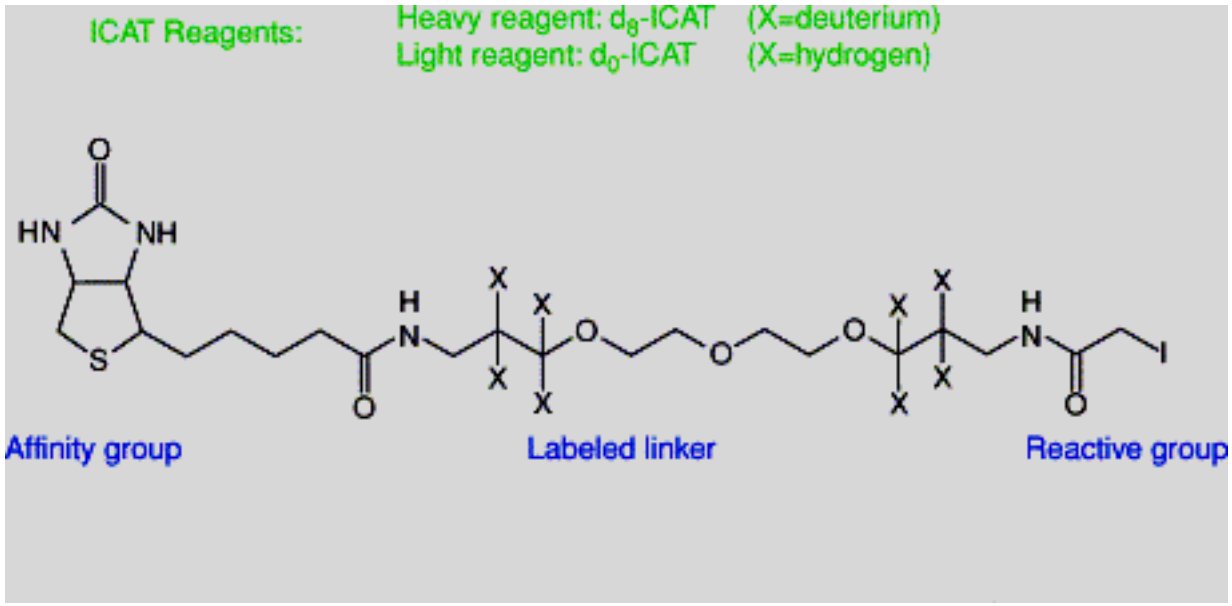
- Solubility—modification has to be low
- Migration rate—modified proteins migrate differently on the gel.
- Reproducibility.
- Sensitivity—the dye by itself is not that sensitive, many proteins not modified. Need dye visualization.
- Resolution—gel resolution is bad.

Modification with deuterated acrylamide

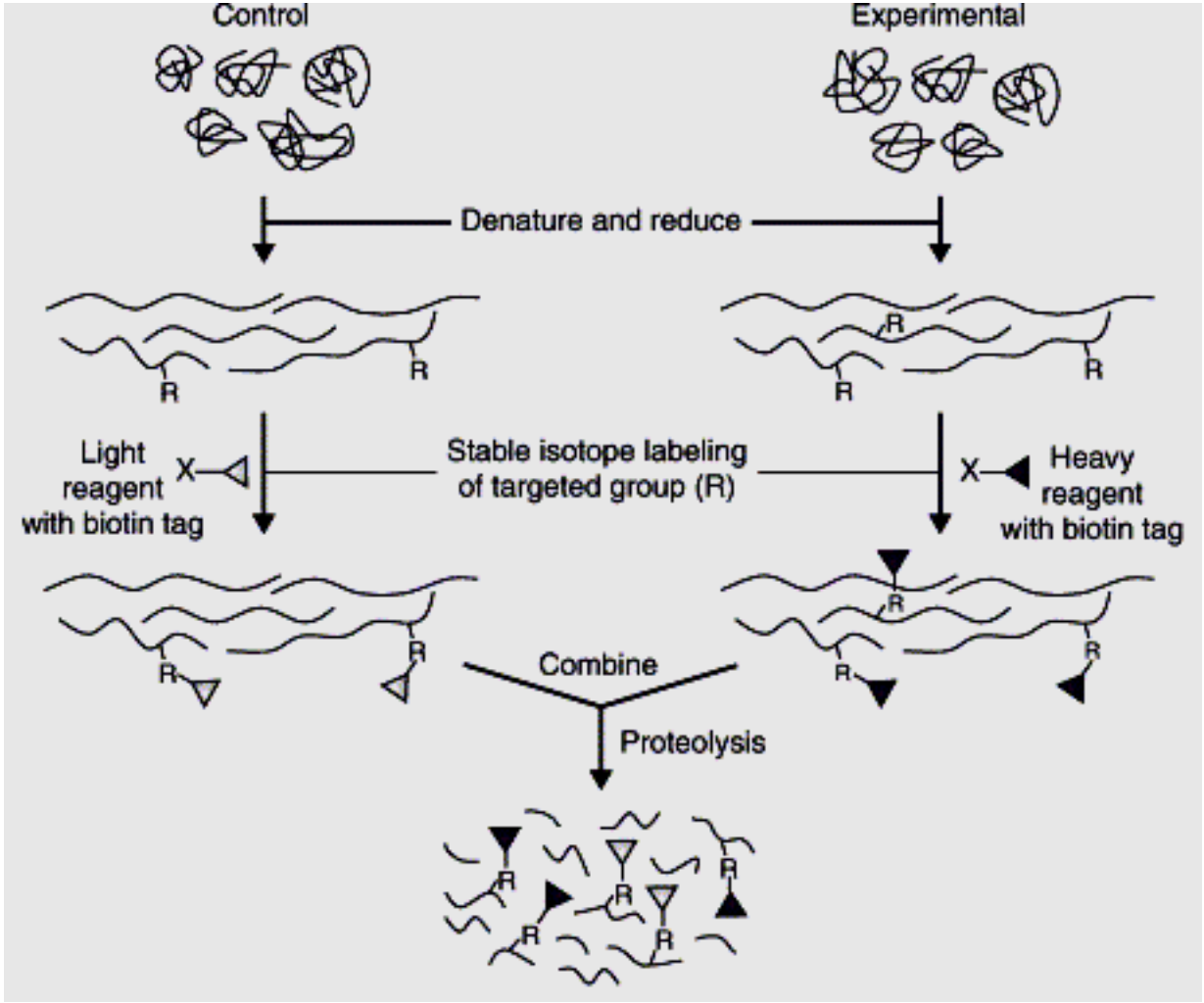
Sechi *Rapid Commun Mass Spectrom* 2002, 16: 1416-1424

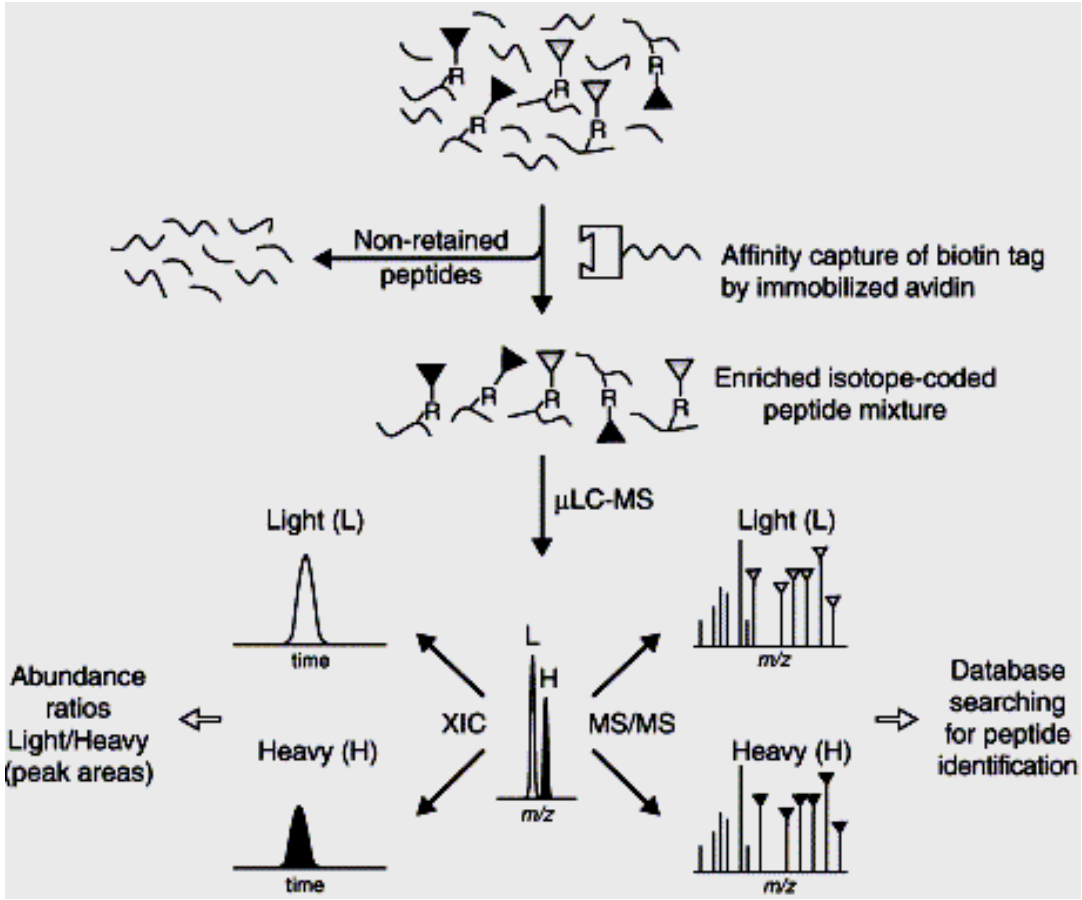


Isotope-coded affinity tags (ICAT)

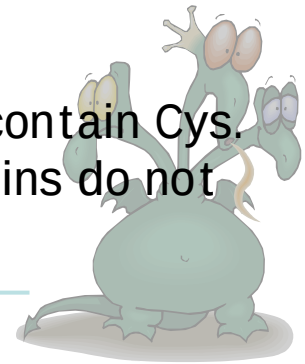


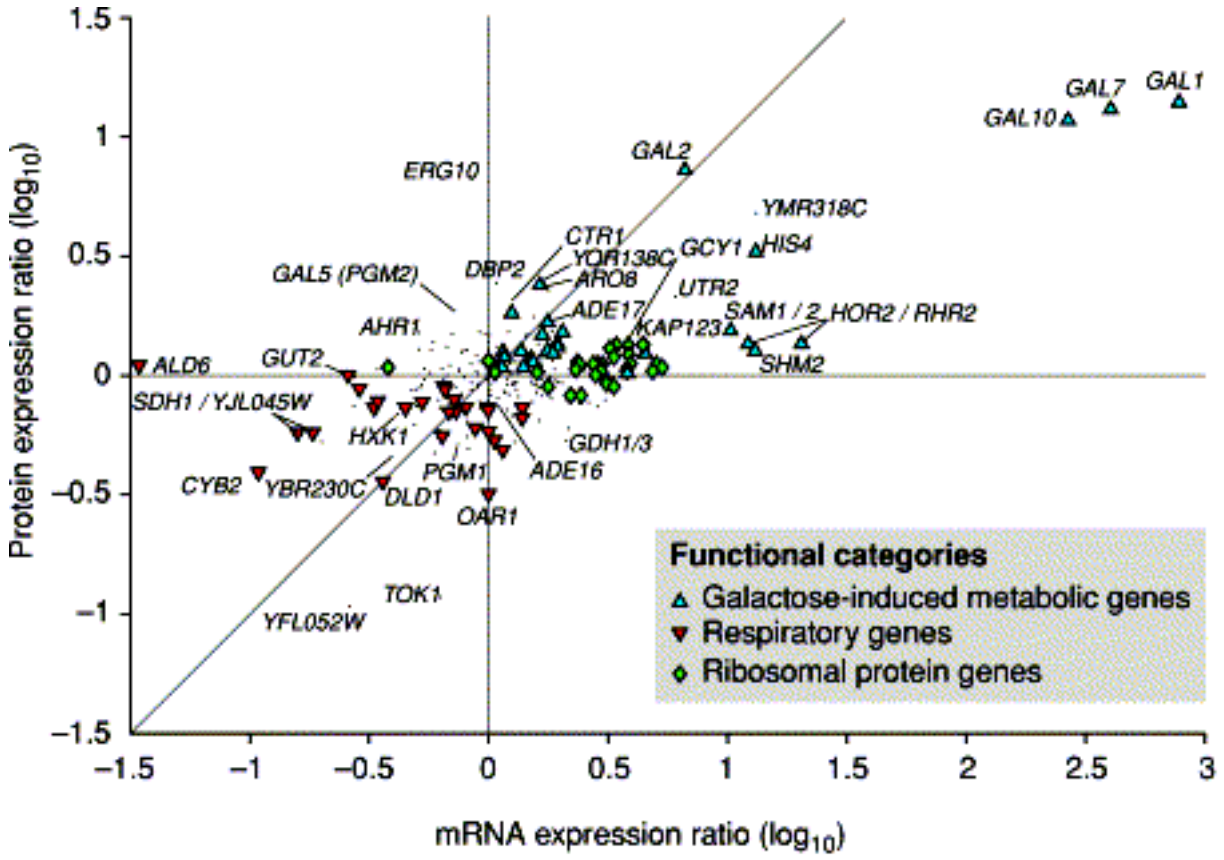
Couples with MudPit
Looking at a subset of proteins


Control
Experimental
Denature and reduce
R
Light
reagent
with biotin tag
X
Stable isotope labeling
of targeted group (R)
X
Heavy
reagent
with biotin tag
Combine
Proteolysis



Yeast genome, ~350,000 tryptic peptides and ~30,000 contain Cys. Provide 92% coverage of the proteome, 8% of the proteins do not have cysteine.





Ideker et al. *Science* 2001, 292: 929-934

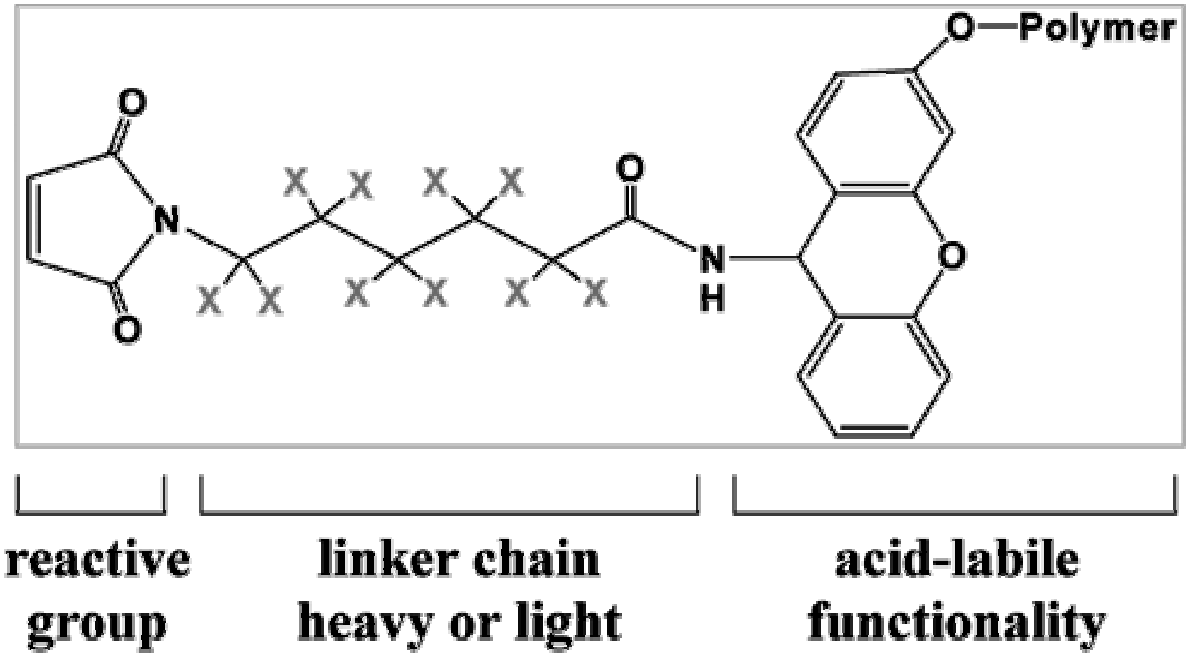
Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network

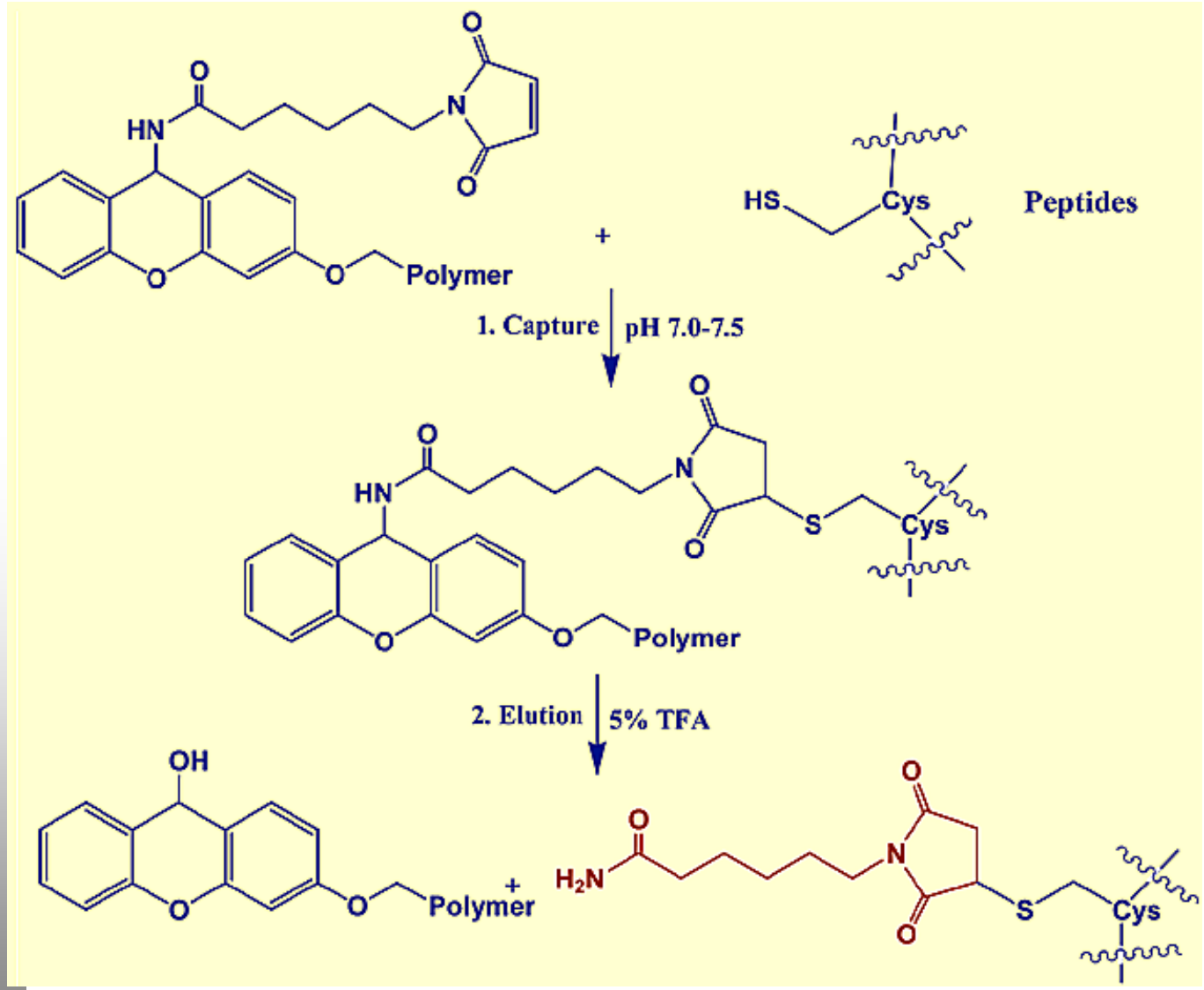
Problems:

- Non-specific binding of non-Cys peptides.
- Large tag mass limits the usefulness for large Cys-peptides.
- Not all proteins have cysteines.
- Difference in retention times between ^{1}H and ^{2}H peptides.
- Fragmentation of label during CID.

Cleavable ICAT reagents

Qiu et al. *Anal Chem* 2002, 74:4969-4979

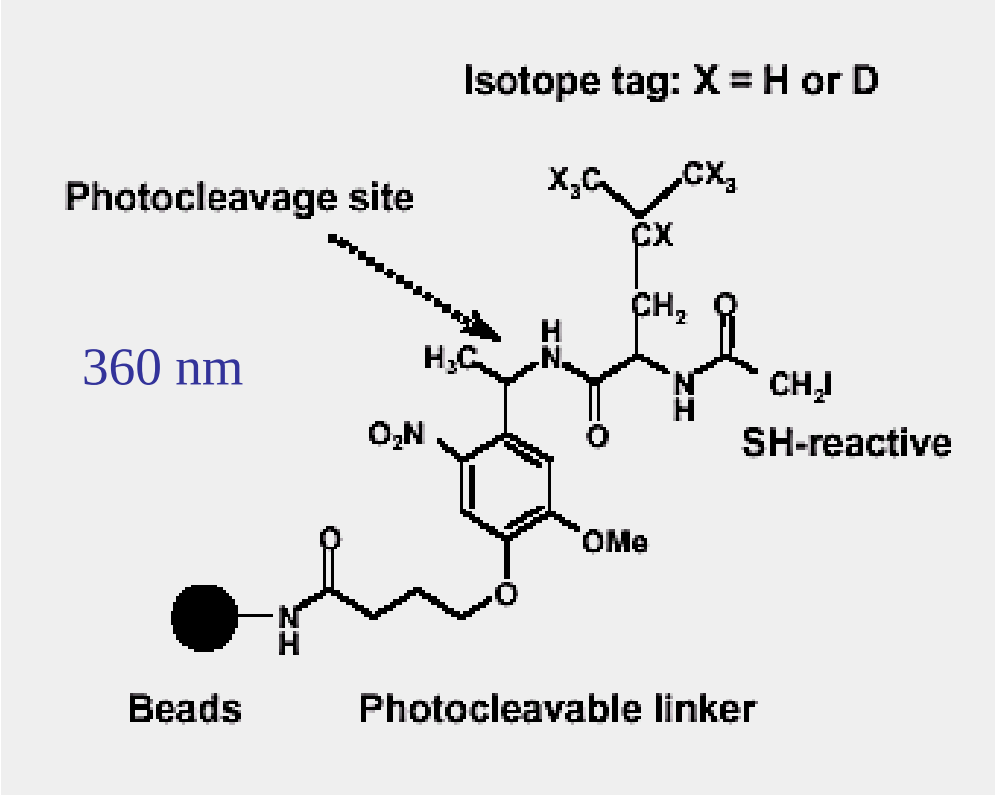



HN
O
N
O
O
O
O
Polymer
+
HS
Cys
Peptides
1. Capture
pH 7.0-7.5
HN
O
N
O
O
S
Cys
O
O
Polymer
2. Elution
5% TFA
OH
O
O
Polymer
+
H2N
O
O
N
O
S
Cys

The ICAT reagent from ABI

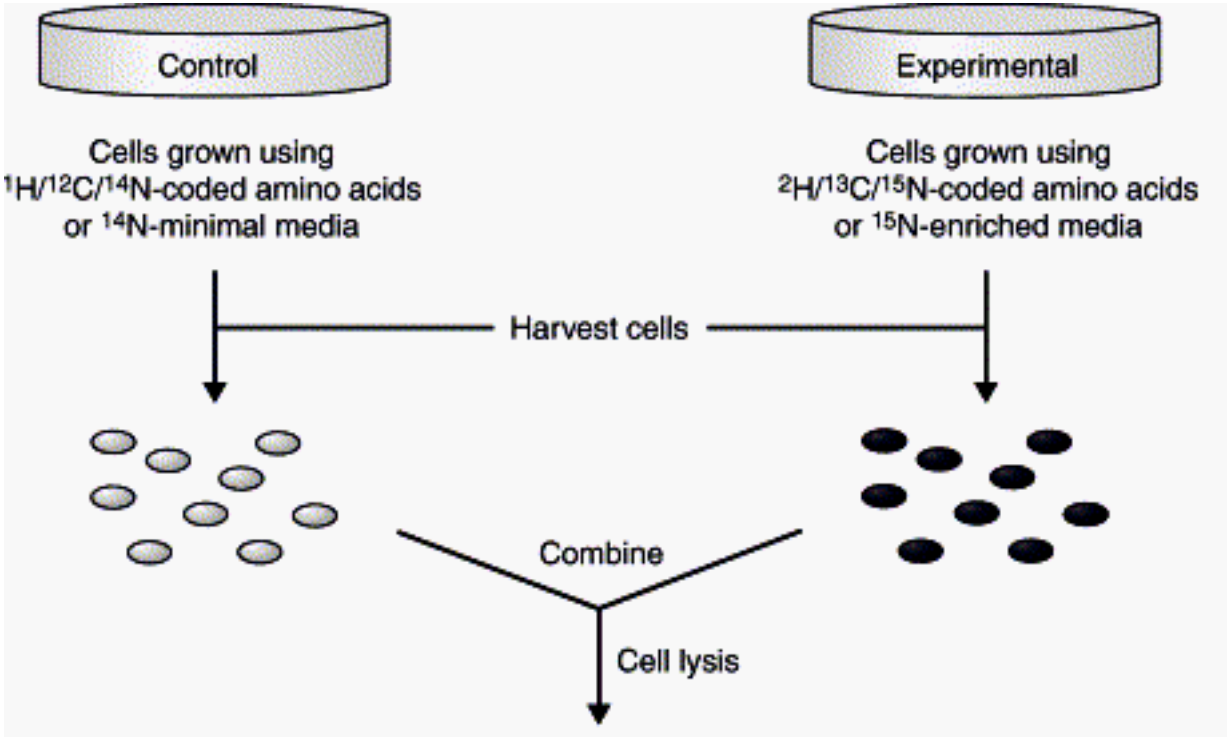
- Acid cleavable biotin group
- 9 Da/labeled cysteine
- Utilize $^{12}C/^{13}C$

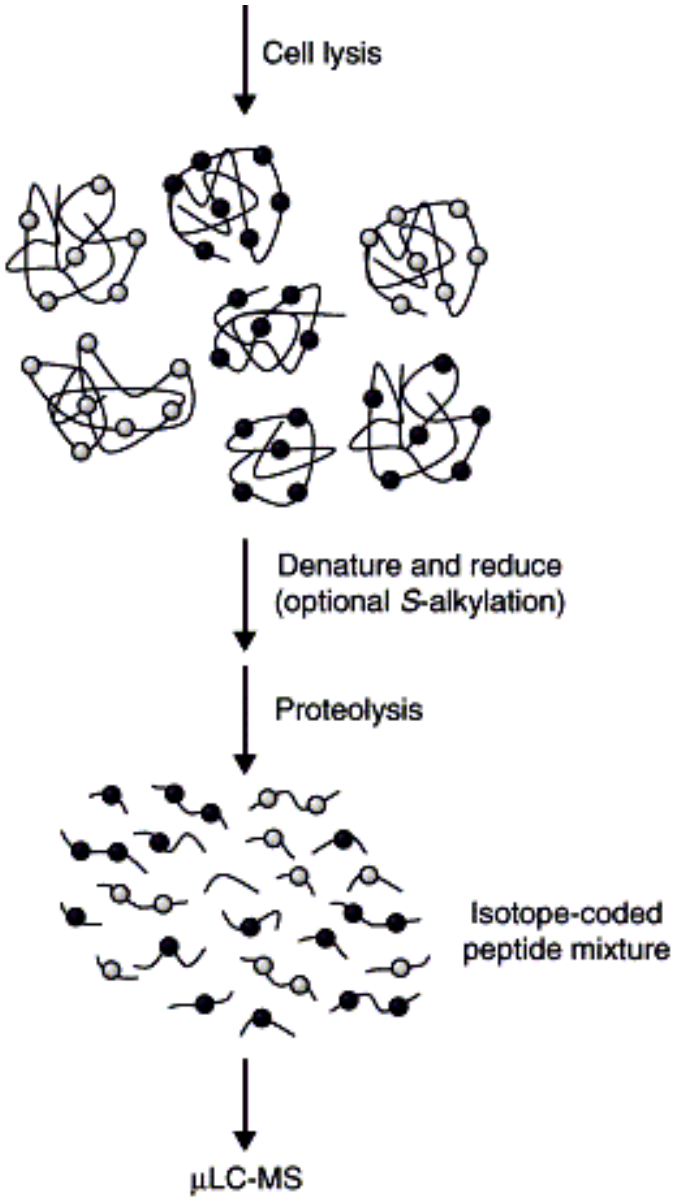
Zhou et al. *Nature Biotech* 2002, 20:512-515



By pass the acid cleavage step

In vivo labeling

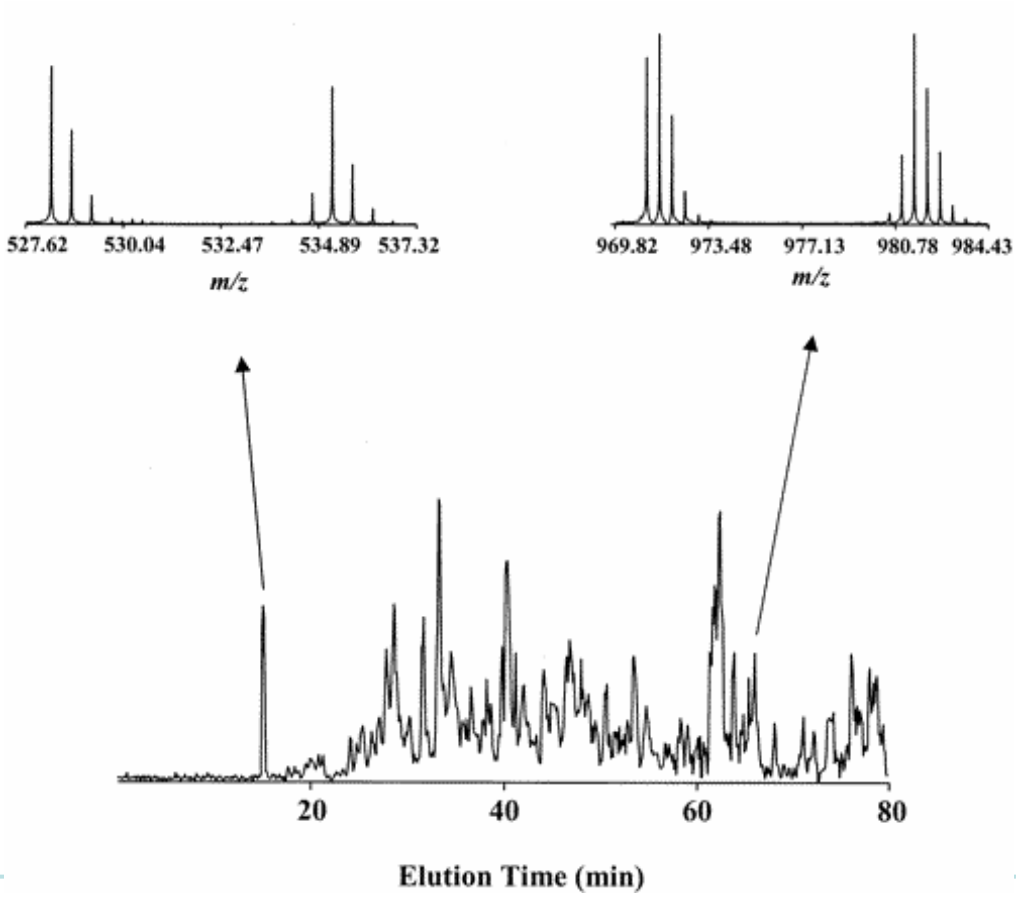



Cell lysis
Denature and reduce
(optional S-alkylation)
Proteolysis
Isotope-coded
peptide mixture
µLC-MS

^{15}N-labeling

Washburn et al. *Anal Chem* 2002, 74: 1650-1657

Conrads et al. *Anal Chem* 2001, 73: 2132-2139



SILAC (stable isotope labeling by amino acids in cell culture)

Leu-d_0 and Leu-d_3

Ong et al. *Mol Cell Proteomics* 2002, 1: 376-386

Leu-d_0 and Leu-d_{10}

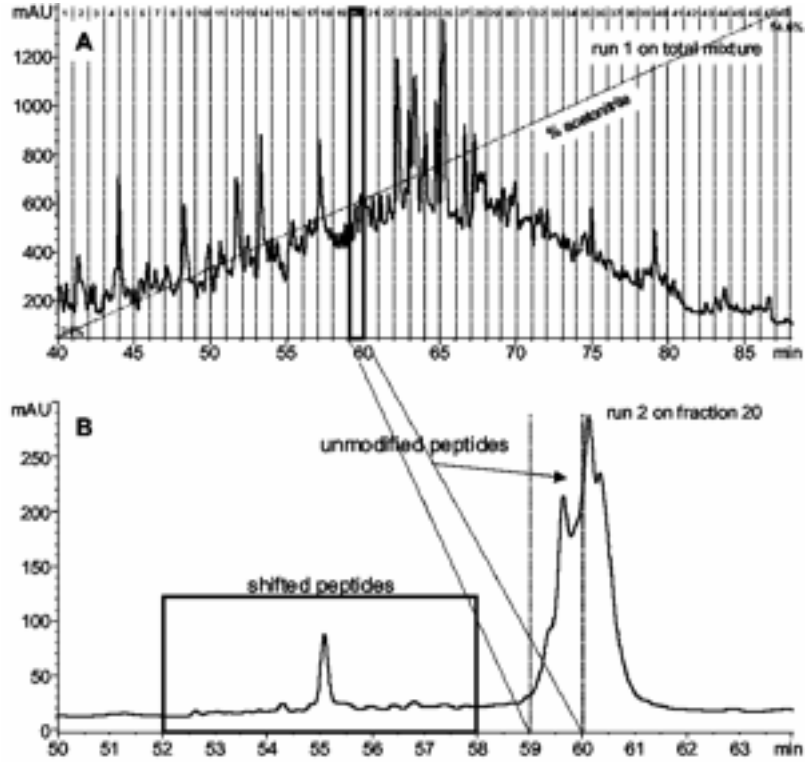
Jiang & English *J Proteome Res* 2002, 1: 345-350

Lys-d_0 and Lys-d_4

Gu et al. *J Am Soc Mass Spectrom* 2003, 14: 1-7

COFRADIC method
(combined fractional diagonal chromatography)

Isolate methionine containing peptides by oxidation



Phosphoproteome

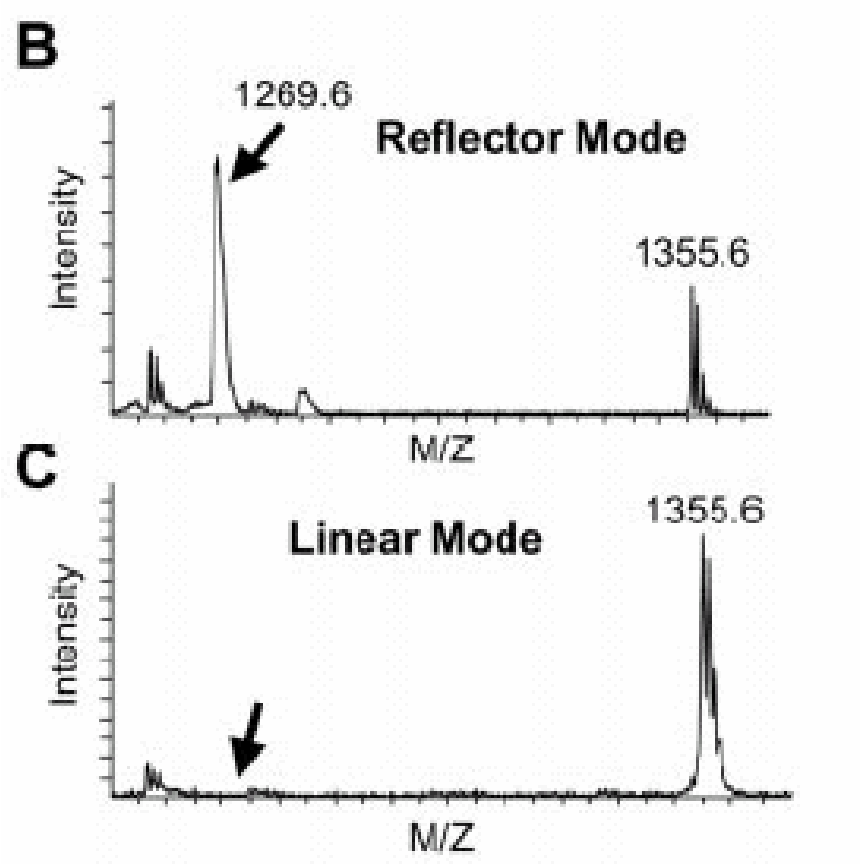
- Estimated 100,000 potential phosphorylation sites
- Less than 2000 are known
- Sub-stoichiometric
- Signal of phosphopeptides vs non-phosphopeptides
- Fractionation/concentration
  - Charges on the peptide, interaction with SS
  - Concentrate peptides on Fe^{3+} and Ga^{3+} IMAC

    Vener et al. *J. Biol. Chem.* 2001, 276: 6959-6966

    Ficarro et al. *Nature Biotech* 2002, 20: 301-305
  - IP of phophoproteins

    Pandey et al. *Proc. Natl. Acad. Sci. USA* 2000, 97: 179-184

J. Biol. Chem. 2001, 276: 6959-6966



Ficarro et al. *Nature Biotech* 2002, 20: 301-305
Jeffrey Shabanowitz/Donald Hunt

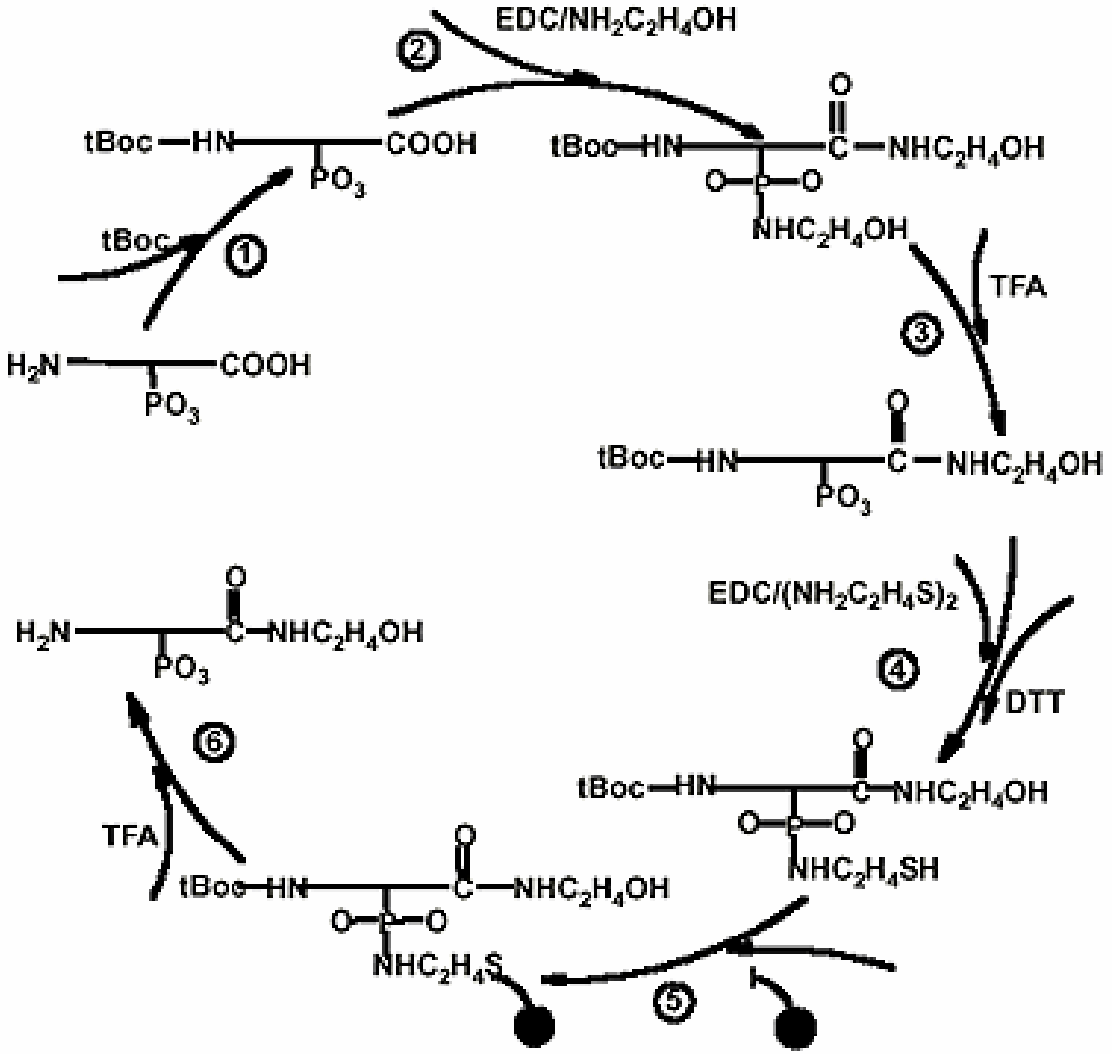
DRVpYIHPF tryptic digests.

Block the carboxyl groups on the peptides.

Chemical tagging

Zhou et al. *Nature Biotech* 2001, 19: 375-378
Aebersold's group

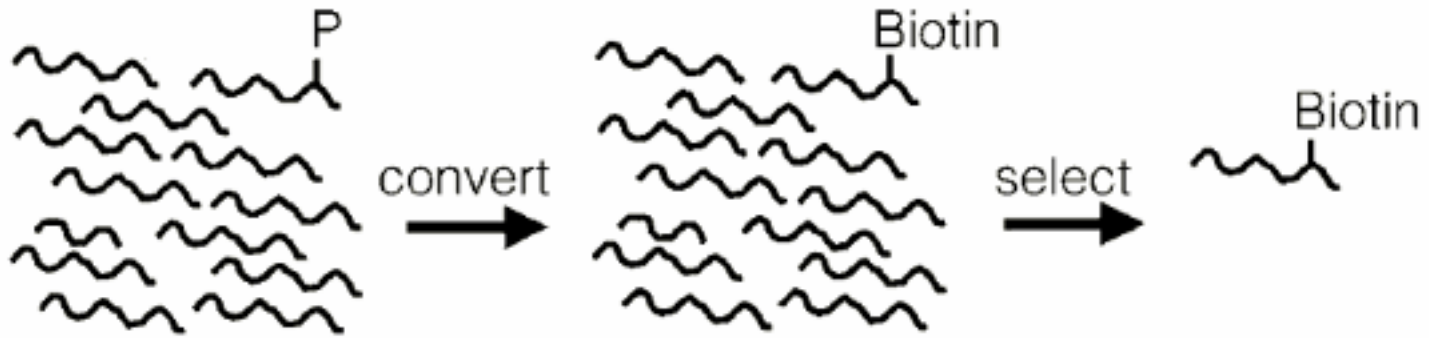
1. t-Boc protection
2. Form amide and phosphoramidate bonds (carbodiimide catalysis)
3. Regenerate phosphate
4. Attach cystamine to regenerated phosphate
5. Capture the peptide with immobilized iodoacetyl groups
6. Recover P-peptide with TFA cleavage

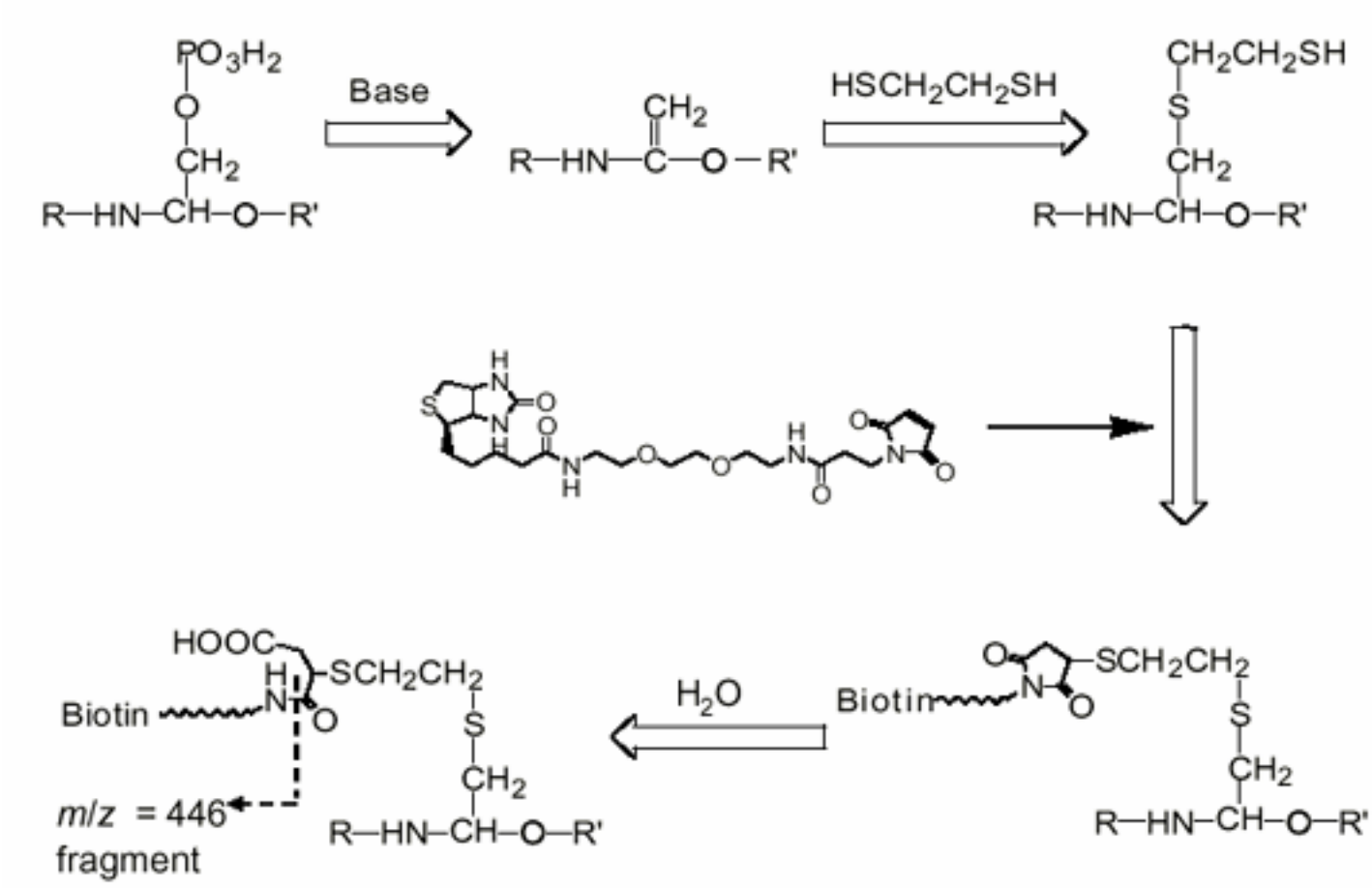

EDC/NH2C2H4OH
tBoc—HN—COOH
PO3
tBoc
H2N—COOH
PO3
tBoc—HN—C—NHC2H4OH
O—P—O
NHC2H4OH
TFA
tBoc—HN—C—NHC2H4OH
PO3
EDC/(NH2C2H4S)2
DTT
tBoc—HN—C—NHC2H4OH
O—P—O
NHC2H4SH
H2N—C—NHC2H4OH
PO3
TFA
tBoc—HN—C—NHC2H4OH
O—P—O
NHC2H4S

Oda et al. *Nature Biotech* 2001, 19: 379-382
Brian Chait's group

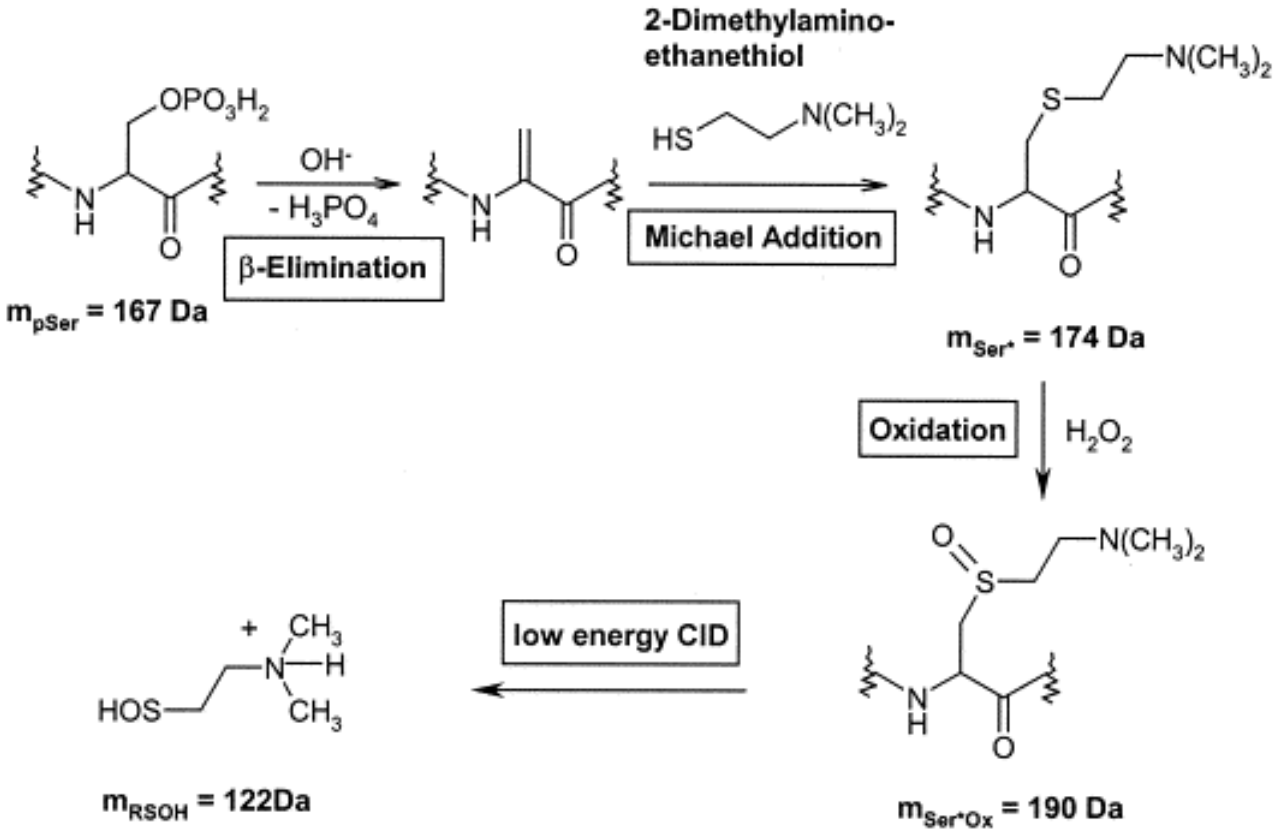
Identify phosphoserine and threonine

β-elimination at pSer and pThr to form dehydroalanyl residues




PO3H2
O
CH2
R–HN–CH–O–R'
Base
CH2
R–HN–C–O–R'
HSCH2CH2SH
CH2CH2SH
S
CH2
R–HN–CH–O–R'
HOOC
H
N
O
SCH2CH2
Biotin
S
CH2
m/z = 446
fragment
R–HN–CH–O–R'
H2O
O
N
O
SCH2CH2
Biotin
S
CH2
R–HN–CH–O–R'

Steen & Mann *J Am Soc Mass Spectrom* 2002, 13: 996-1003



Read:

Reinders & Sickmann

Proteomics 2005, 5, 4052-4061

“State-of-the-art in phosphoproteomics”

Larsen et al.

Mol Cell Proteomics 2005, 4, 873-886

“Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns”