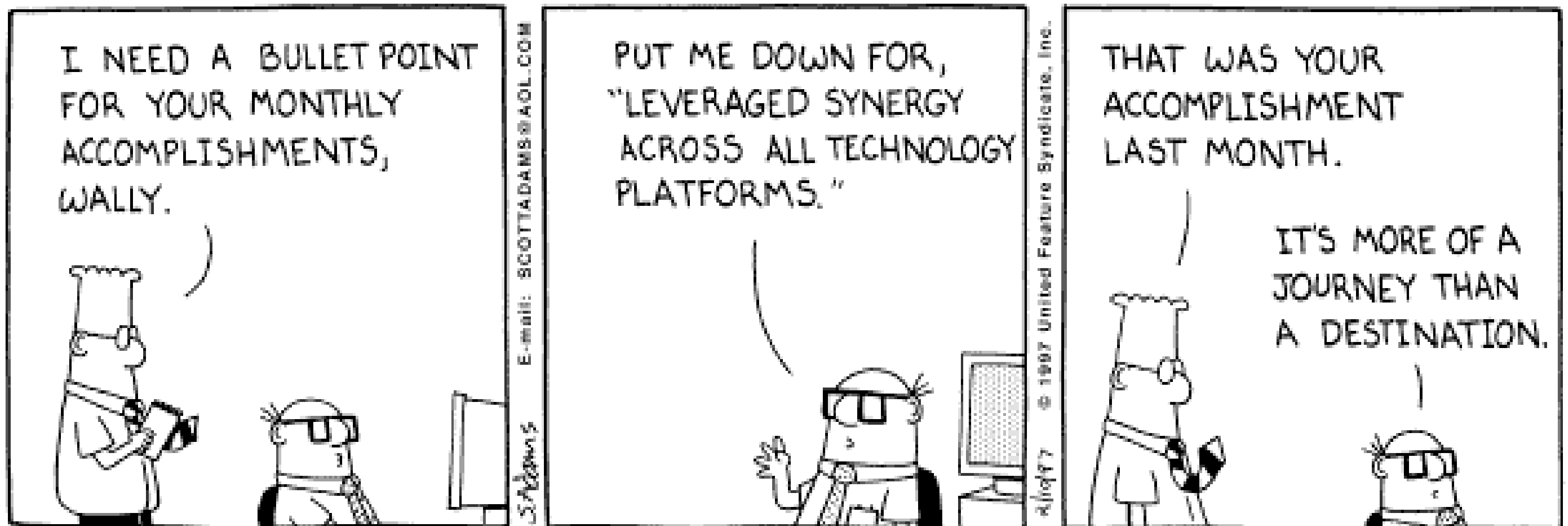
Biomolecular Applications
of LC/MS

LC/MS for biomolecule applications:
"leveraged synergy across many technology platforms"



Copyright © 1997 United Feature Syndicate, Inc.
Redistribution in whole or in part prohibited

Overview

- Biomolecule LC/MS basics
  - mass spectrometry (MS) instrumentation
    - triple quads, ion traps, MS, MS/MS, MS^n
  - electrospray ionization (ESI)
    - mass range, multiple charging
  - Peptide sequencing
    - database searching for protein i.d.
    - De novo sequencing *(a more advanced topic)*
- Methods and Applications
  - Case studies
    - Protein i.d. from gel-separated proteins
    - Applications in biopharmaceuticals
  - Automation strategies
    - Sample cleanup
    - On-line digestion
    - On-line interaction screening
    - Data-handling strategies

Information Obtainable via ESI/LC/MS^n

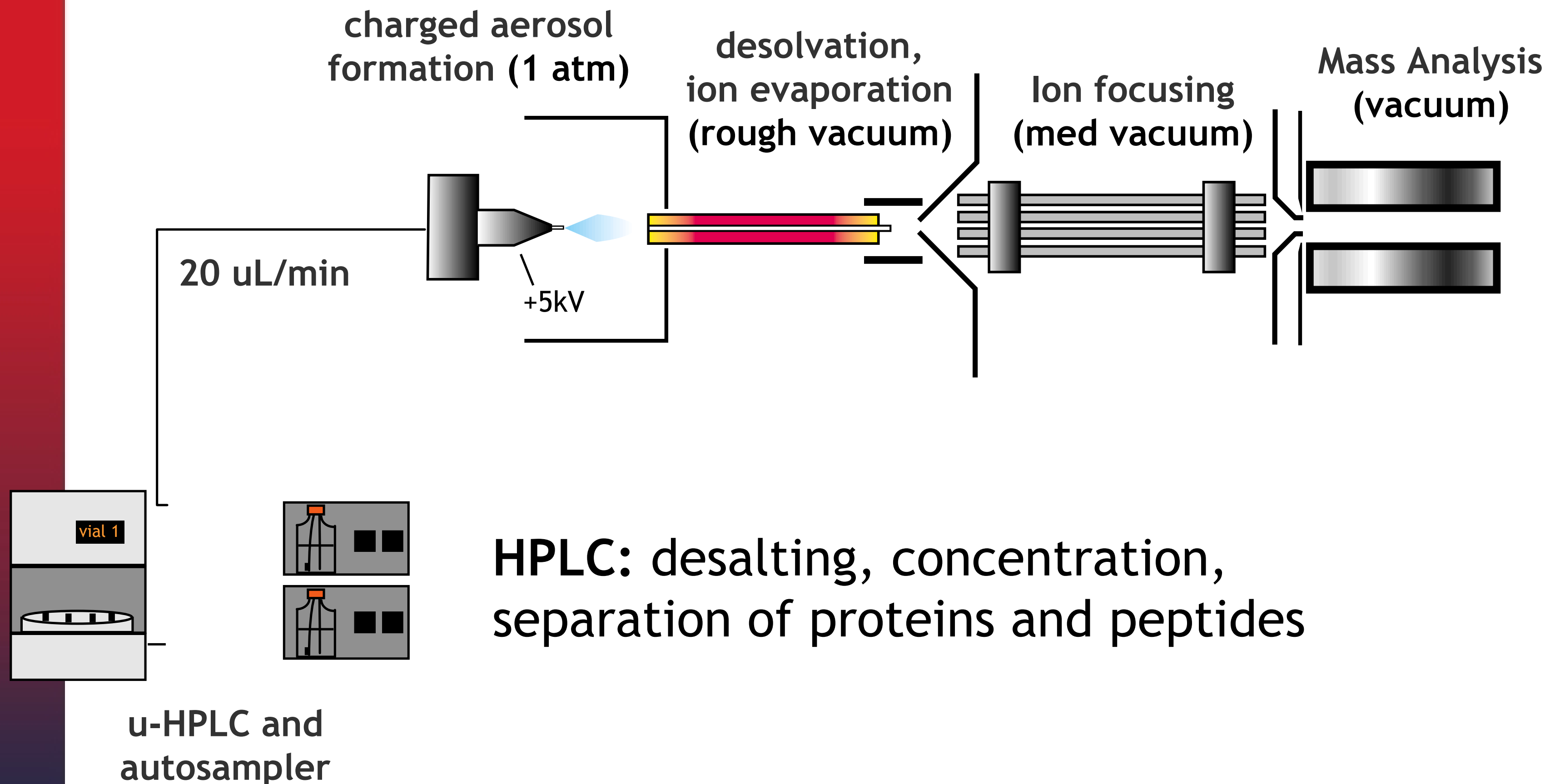
- Molecular Weight
  - confirmation of sequence
  - screening for post-translational or other covalent modifications
  - heterogeneity/impurity profiling
- Structural Information (via MS/MS or MS^n)
  - sequence confirmation
  - protein identification from database searching of fragment ions
  - disulfide bond mapping
  - determination of sites of modification or mutation
- Binding/Interaction
  - on-line affinity selection or affinity chromatography

Challenges/limitations of LC/MS for Biomolecular Analysis

- ESI/MS of intact proteins can be difficult if protein is too heterogeneous
- most protein i.d. requires enzyme digestion first
- detergents interfere with LC/MS analysis and must be removed which is non-trivial
- low tolerance of ESI to salt and 100% aqueous solns makes analysis of proteins in their native state difficult
- Information and data overload!

LC/Electrospray Ionization/MS

Electrospray ionization (ESI): separation of ions from bulk solution so that they may be analyzed in the gas phase (mass spectrometry)

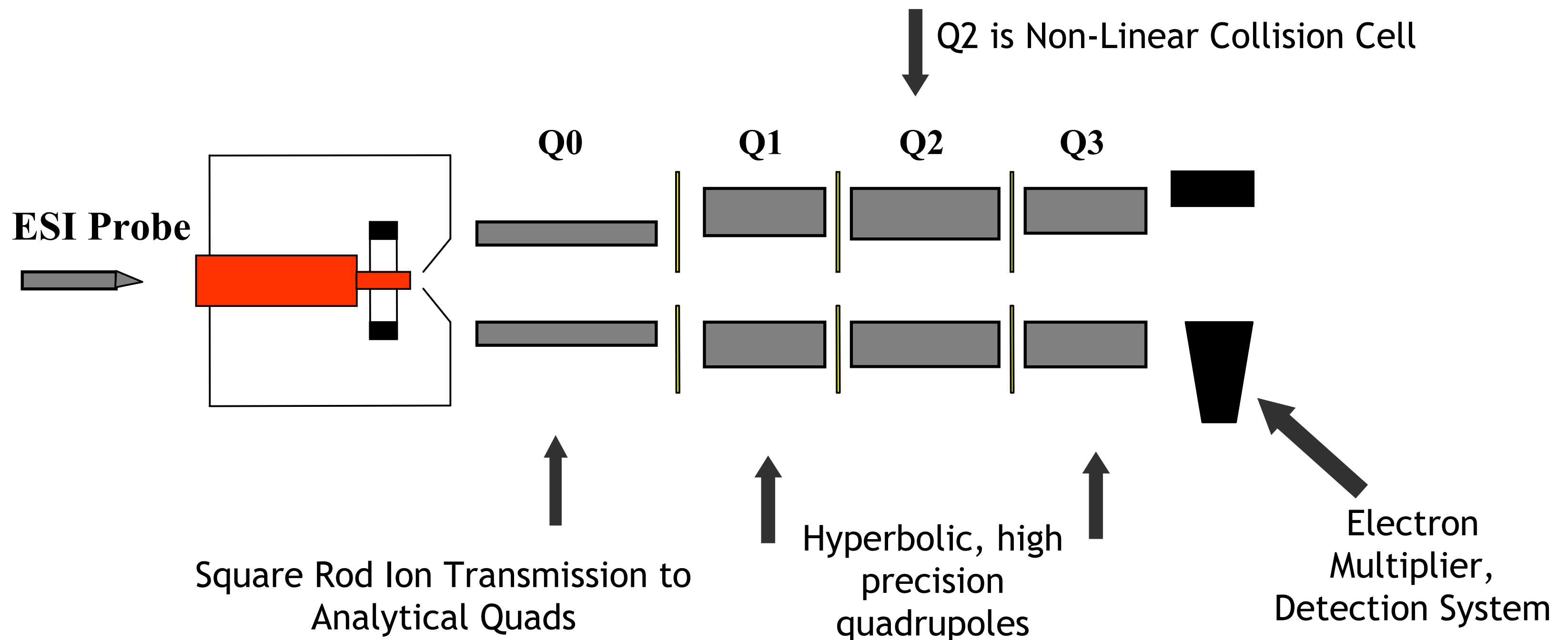


HPLC: desalting, concentration, separation of proteins and peptides

Schematic Diagram of a Triple Quadrupole Mass Spectrometer

A "tandem-in-space" mass spectrometer

different stages of mass analysis performed in discrete regions (quads 1 or 3)

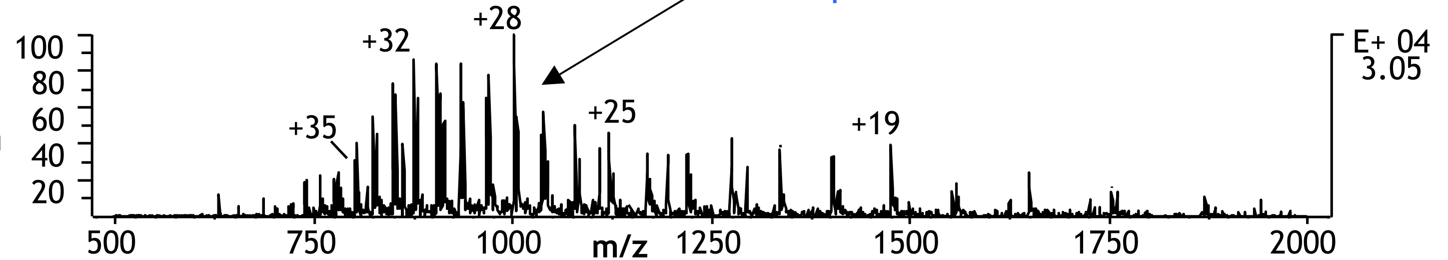


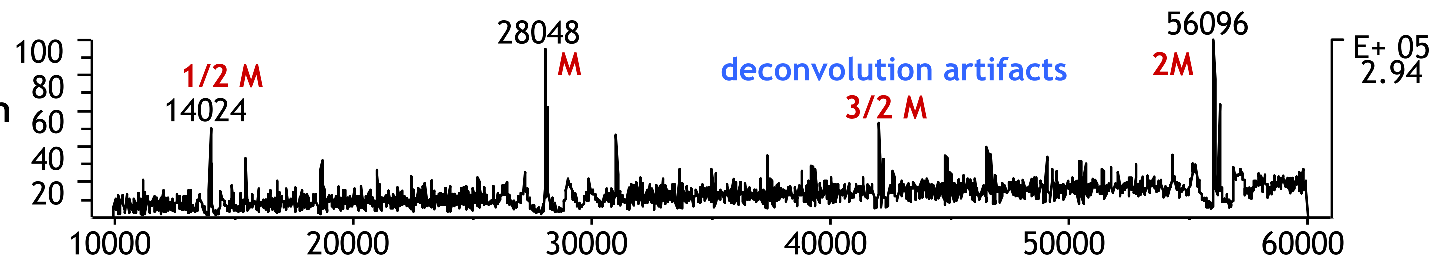
strengths: great general purpose system MS, MS/MS, neutral loss and parent scans useful for mixture screening, best instrument for intact protein analysis, no low mass cut-off in MS/MS mode like the ion trap, data-dependent scanning

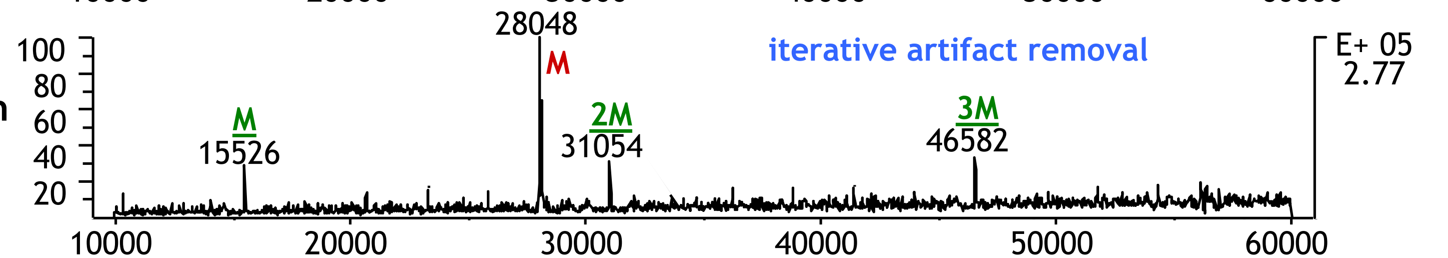
limitations: less sensitive than ion trap in the full scan MS/MS mode

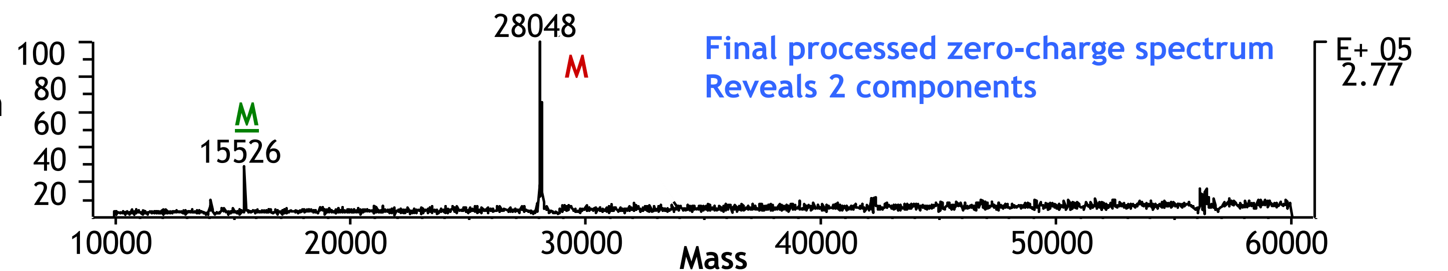
Determination of Protein MWs using Deconvolution

HCMV Protease (MW 28049)

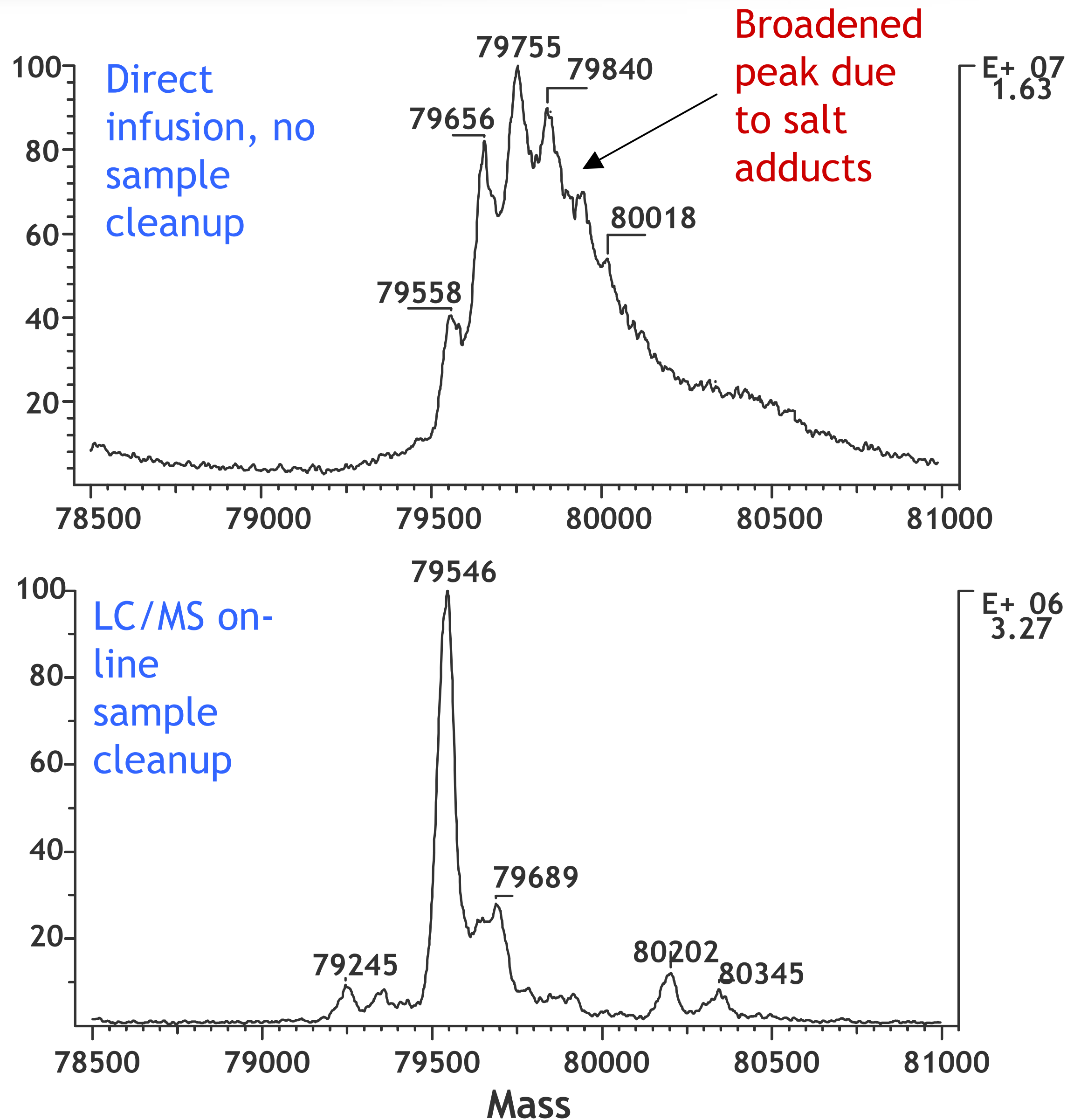




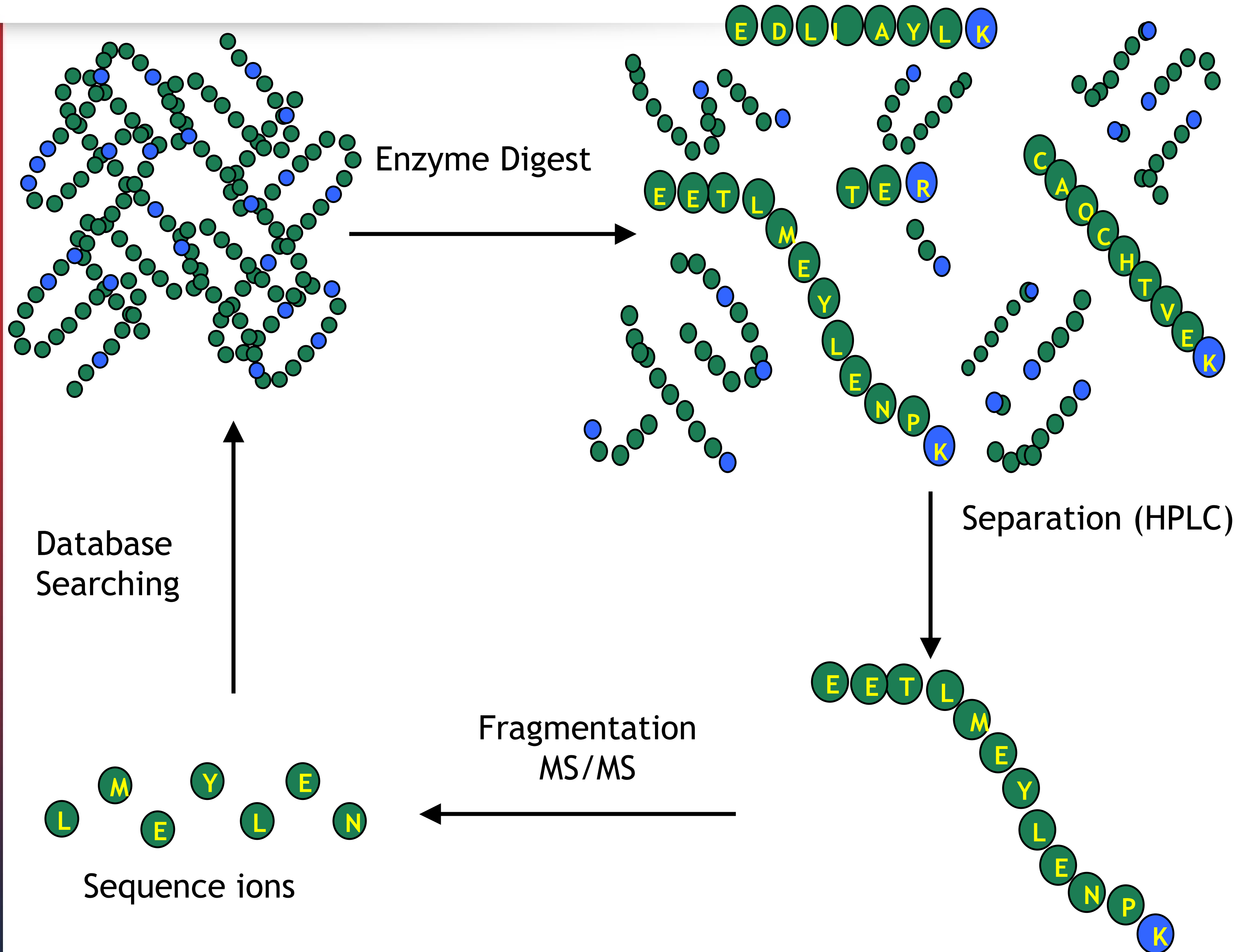




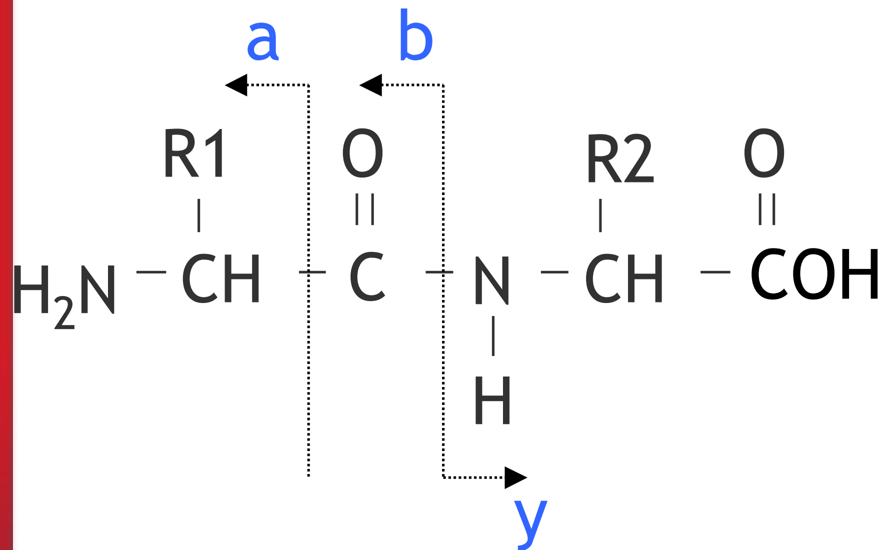
Importance of protein sample cleanup by LC/MS: analysis of 80 kDa human transferrin



Proteins require enzymatic cleavage before sequencing can occur by MS



Fragmentation in MS/MS produces sequence-specific fragment ions



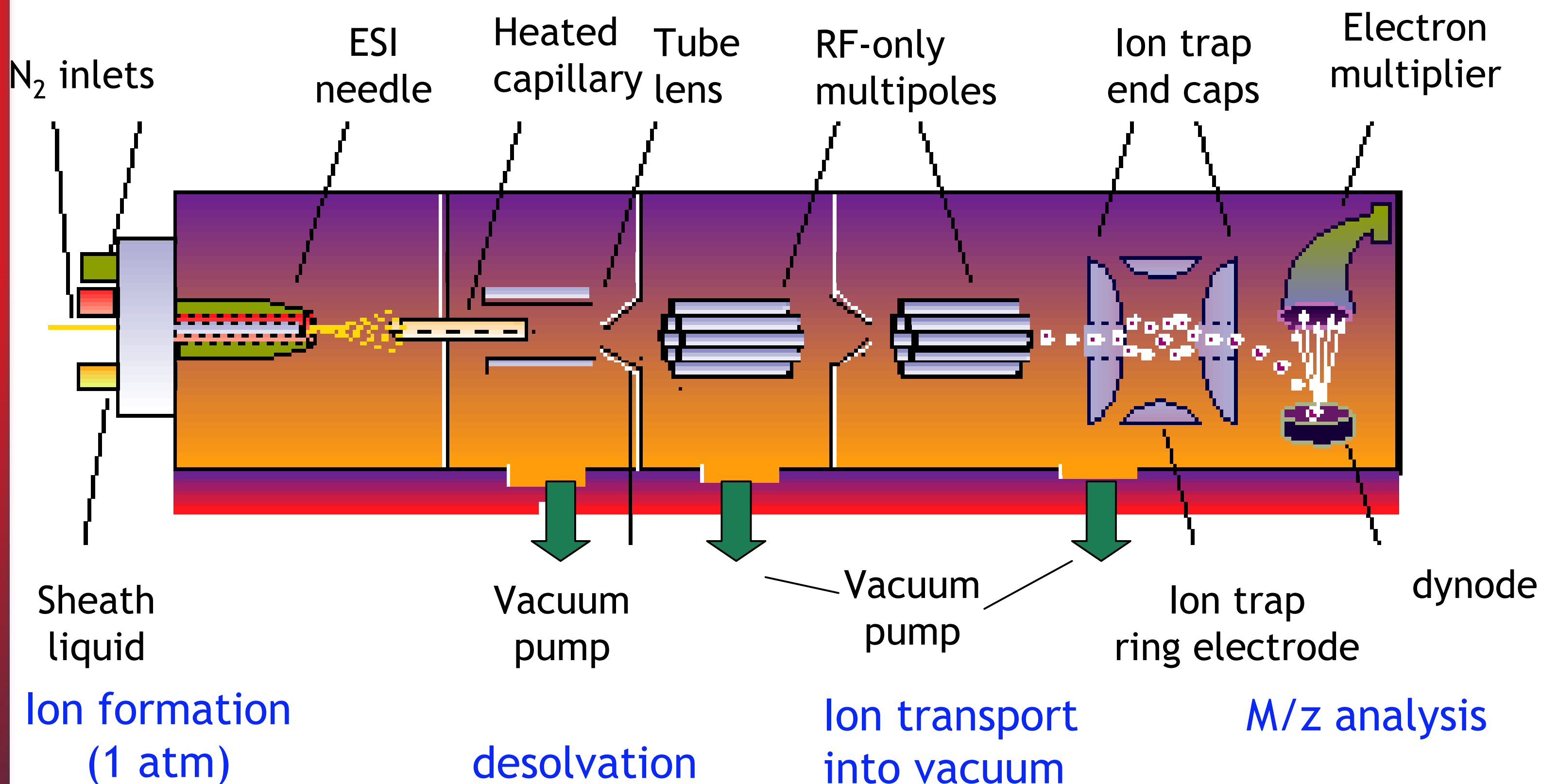
- Use database search routines to identify or profile known proteins
- *De novo* sequencing requires interpretation of spectra either manually or by computer algorithm

- peptides fragment along the amide backbone to produce sequence specific fragment ions
- different patterns of sequence ions are produced from the different masses of the side chain groups
- ions containing the N-terminus are typically of type 'a' or 'b'
- ions containing the C-terminus are known as 'y' ions
- tryptic peptides predominantly produce ions of type 'y' due to the presence of Arg or Lys at the C-terminus

Schematic Diagram of the Finnigan LCQ Ion Trap MS

A "tandem-in-time" mass spectrometer

different stages of mass analysis performed in the same physical space at discrete points in time



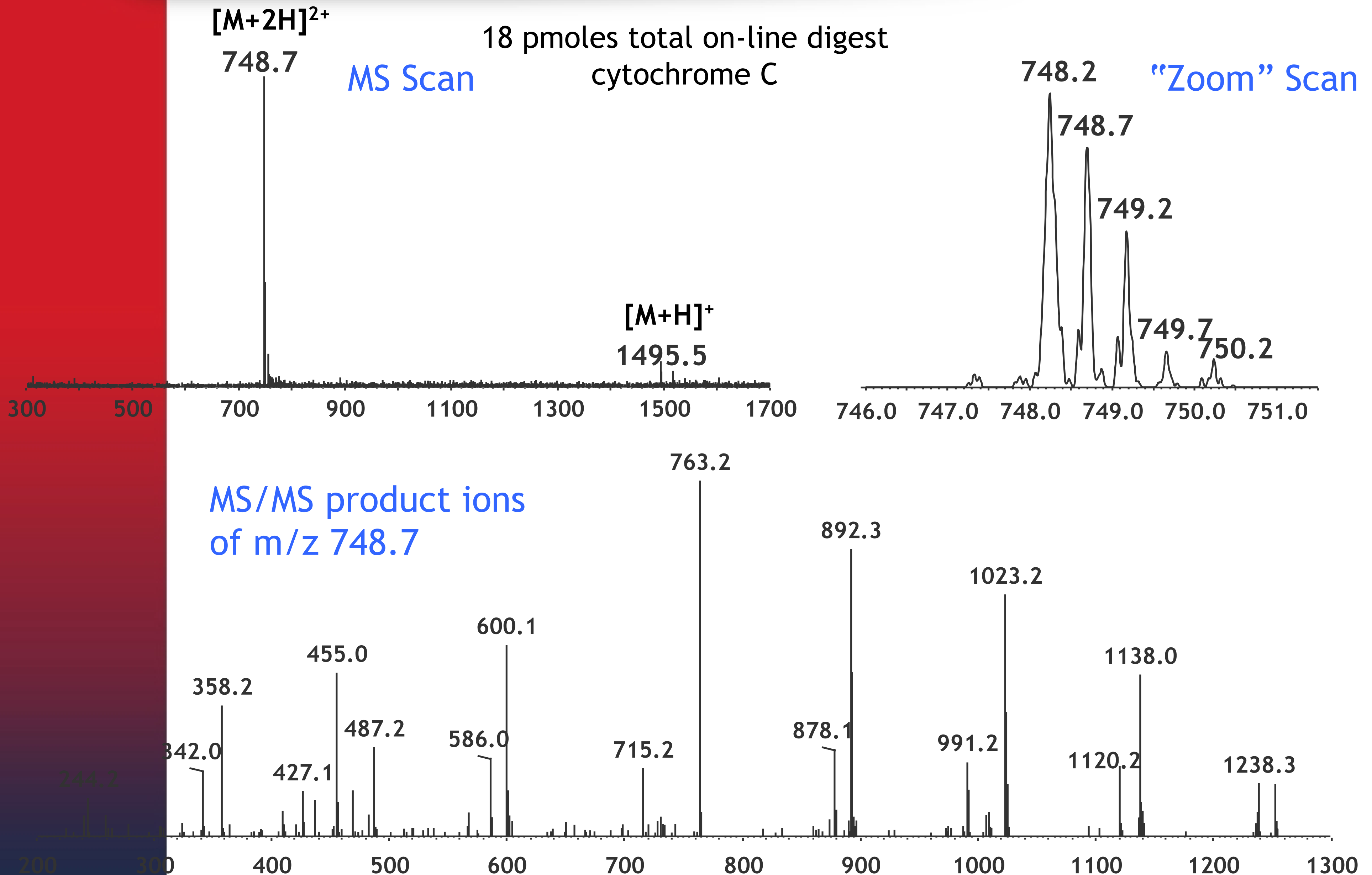
strengths: compact, MS^n, **excellent full scan MS/MS sensitivity**, excellent data dependency, the ultimate peptide mapping system

limitations: limited storage of ions, 1/3 of low mass range lost in MS/MS modes

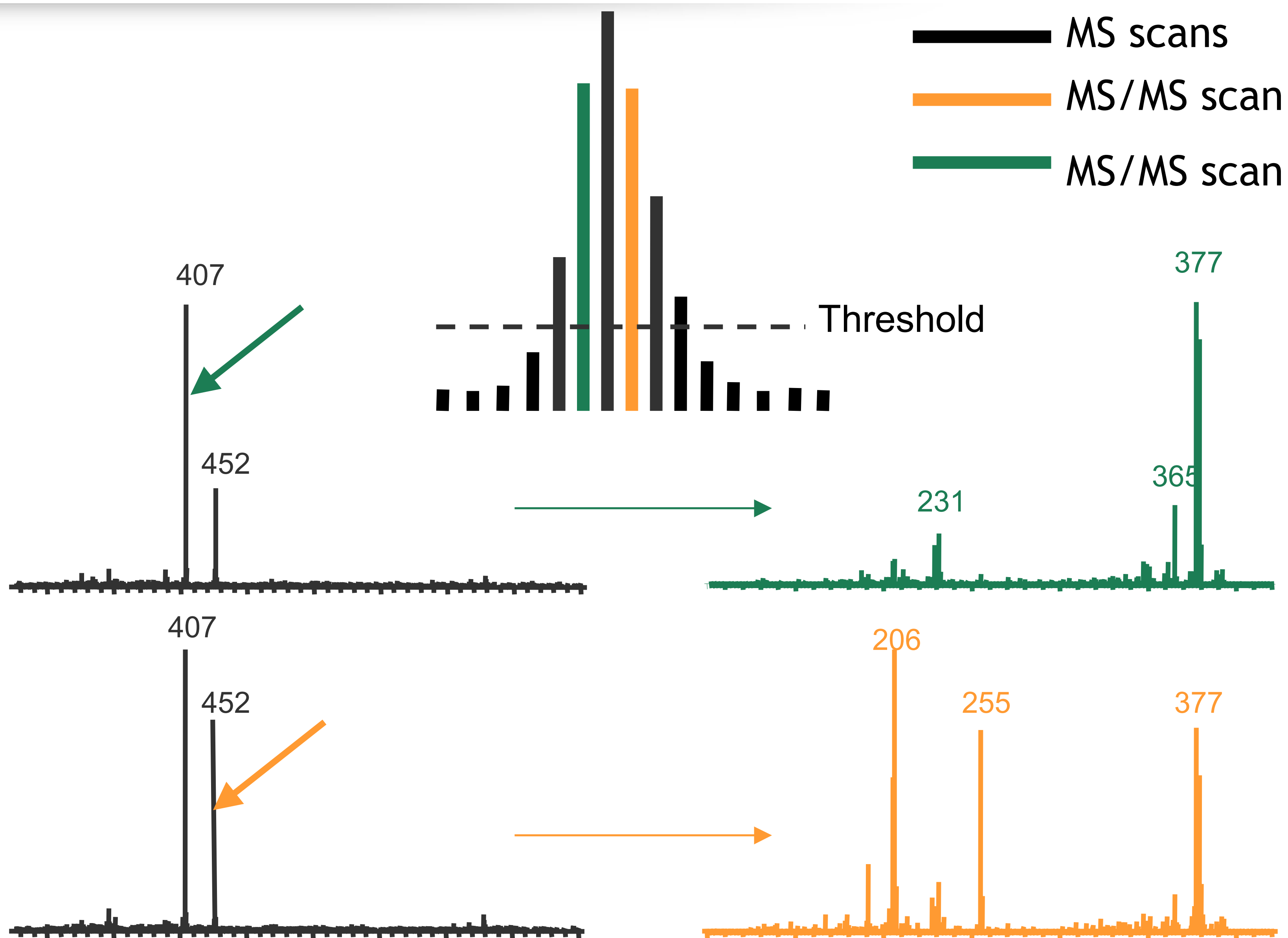
Ion trap scanning terminology for bio-LC/MS applications

- Data-dependent scanning
  - Having the instrument switch scan modes once a signal threshold has been exceeded
  - A key advantage of the Finnigan ion trap systems
  - Maximizes information content in a single LC/MS^n run
- LCQ "Triple Play" – a data-dependent scan type
  - Full scan (find signals for MS/MS)
  - Zoom scan (determine charge state to calculate MW)
  - Full MS/MS scan (used to delineate peptide sequence, via SEQUEST database search or *de novo* methods
- Dynamic Exclusion – enhanced data-dependency
  - Ion traps yield excellent MS/MS spectra with a single scan
  - Therefore, why waste instrument time collecting additional spectra on the same signals?
  - Dynamic exclusion – ignore masses (for a period) whose MS/MS spectra have already been collected
  - Allows "drilling down" further into the data
  - Applications: complex mixtures of co-eluting peaks, e.g., complex peptide maps

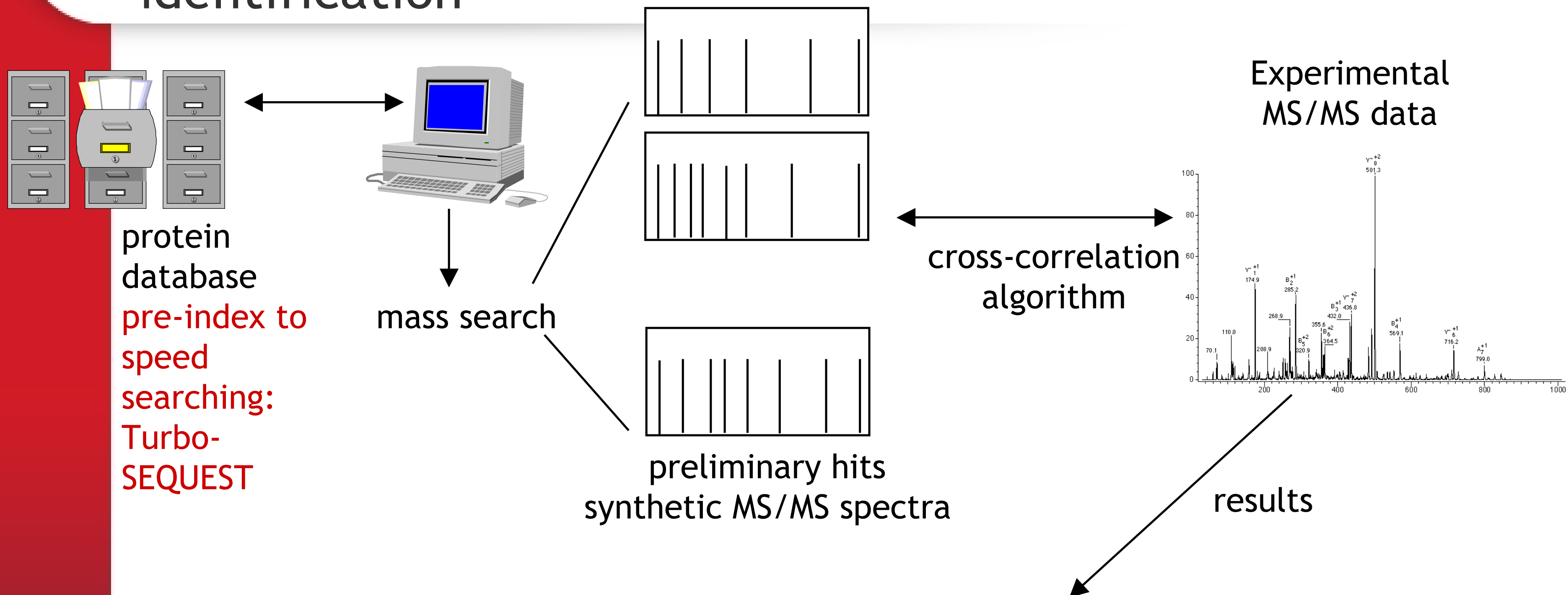
Example of LCQ "triple play" data-dependent analysis



The "Dynamic Exclusion" Experiment



SEQUEST database cross-correlation for rapid protein identification



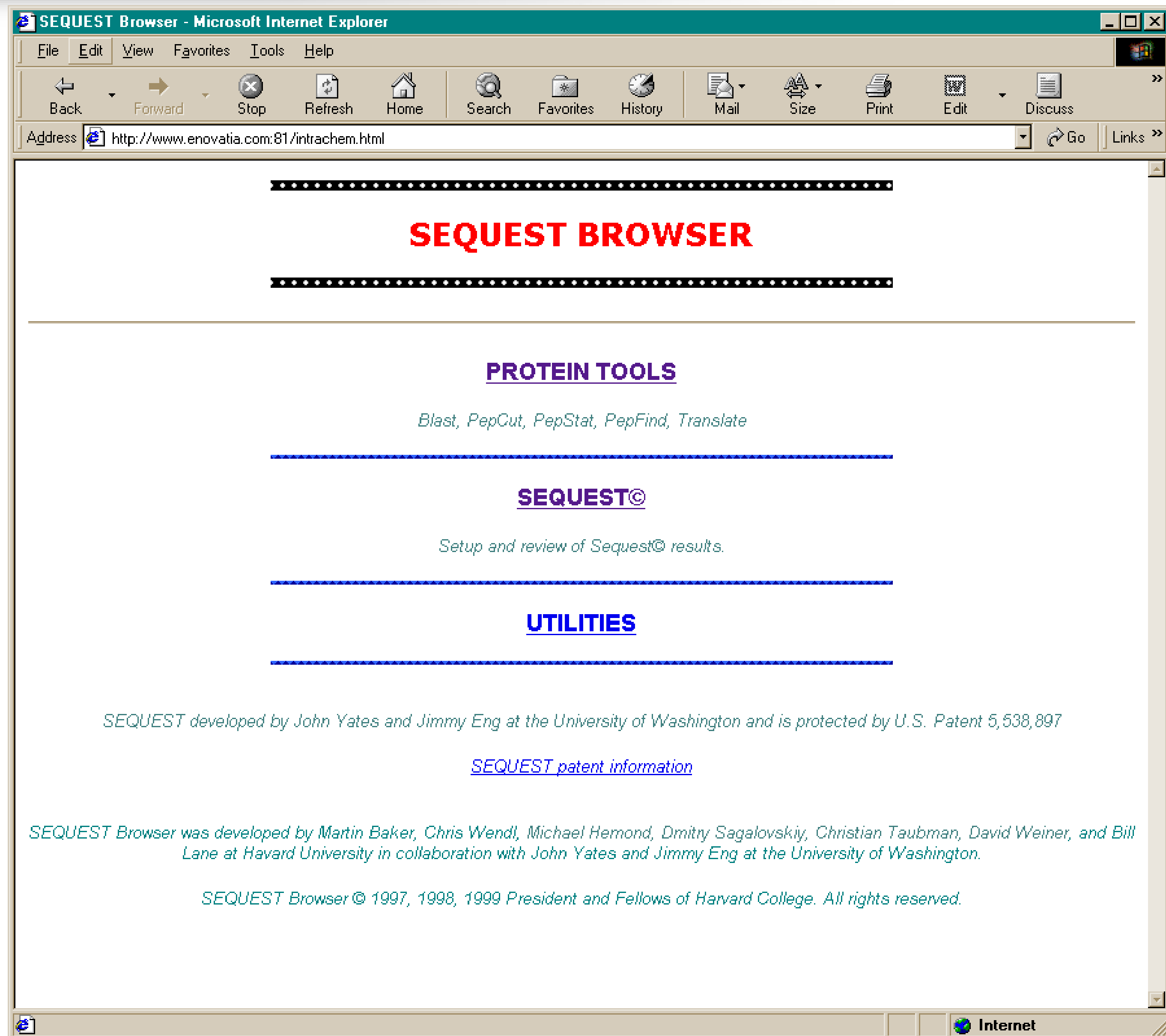
Database Search Results

Database:genpept.fsa   Mass:1001.00   Charge:2
File:ERFHASVR.dat   Scan:Off-line centroid

No.	Rank	Corr.	Protein Information	Mult	Ions	NCBI#	Subsequence
1)	1	1.000	(M88011) glucokinase [Homo sapiens]	+ 3	13/14	179427	ERFHASVR
2)	2	0.976	(U61948) similar to the non-receptor class of pr		10/12	1397280	RKRKEKR
3)	2	0.976	(U51032) D9651.5 gene product [Saccharomyces cer	+ 1	10/12	1230662	RKRQEKR
4)	3	0.964	(U08315) calnexin homolog [Arabidopsis thaliana]		10/12	473878	RKRQTRR
5)	4	0.955	(X78823) phenoxybenzoate dioxygenase [Pseudomona		12/14	473250	GVIYHAWR
6)	5	0.940	(Z68001) C11G10.2 [Caenorhabditis elegans]		10/12	1070045	RERQKER
7)	5	0.940	(U09820) helicase II [Homo sapiens]		10/12	606833	RERKQER
8)	6	0.938	(X07384) GLI protein (AA 1-1106) [Homo sapiens]		14/18	31768	ASDPAQAADR
9)	7	0.932	(U19151) putative reverse transcriptase; contain		10/12	624681	RKREEQR
10)	7	0.932	(Z46259) NO388 gene product [Saccharomyces cerev	+ 1	10/12	633674	RKREEKR

see Yates et al., *J. Am. Soc. Mass Spectrom.* 1994, **5**, 976-989

SEQUEST can be run directly from a browser



SEQUEST Summary screen

Sequest Summary - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Size Print Edit Discuss

Address http://www.enovatia.com:81/cgi-bin/runsummary.pl?directory=mhailjay2&sort=consensus Go Links

Sequest Summary

Setup CreateDTA VuDTA RunSequest Status Summary Utilities Home

Sample: Hail, M. (JAY2) JAY2 meh
Datafiles: jay201 (06/02/2000-06/05/2000)
Intensity: Full Zoom MS2
Db: owl (01/10/2000) and contaminants_and_integrin nr
Dir: mhailjay2
Diff Mods: 57 C 16 Mdiff_search_options = N
Inspector Vi
Enz: Trypsin

Consensus Chron Max rank: 3 Max list: 5 Controls: Show DTA VCR

#	TIC	File	z	dM	MH+	Xcorr	dCn	Sp	RSp	Ions	Ref	()	Sequence

Pull to Top

A owl|P05556|ITB1_HUMAN 98 15|4.3e7|70% (9,1,0,1,0,0) (2 3 6 7 8 9 13 14 16 18 21 24 25 31, 27, x, 20, x, x)
*FIBRONECTIN RECEPTOR BETA SUBUNIT PRECURSOR (INTEGRIN BETA-1)

#	TIC	File	z	dM	MH+	Xcorr	dCn	Sp	RSp	Ions	Ref	()	Sequence
8	8.9e5	0438	3	-1.1	1846.1	3.82	0.25	1164	1	26/60	owl|p05556|itb1		(R) DKLPQPVQPDPVSHCK
6	8.0e5	0432	2	0.1	1536.4	3.60	0.30	1298	1	16/20	owl|p53712|itb1	+6	(K) FCECDNFNCDR
25	1.2e7	0575	2	0.3	1091.3	3.59	0.33	1038	1	15/18	owl|p53712|itb1	+2	(K) GEVFNELVGK
31	2.0e6	0690	2	-0.4	1267.0	3.29	0.39	1353	1	17/20	owl|p53712|itb1	+4	(K) SLGTDLMNEMR
18	1.5e6	0531	2	0.2	1282.4	3.26	0.35	1309	1	17/20	owl|p53712|itb1	+4	(K) SLGTDLMNEM*R
16	5.1e6	0511	2	-0.7	1223.3	3.22	0.36	550	1	12/18	owl|p12607|itb0	+10	(K) WDTGENPIYK
3	2.2e6	0420	2	0.2	1298.4	3.01	0.33	1537	1	17/20	owl|p53712|itb1	+4	(K) SLGTDLM*NEM*R
14	5.5e6	0482	2	-0.5	1097.1	2.91	0.31	768	1	14/18	owl|p05556|itb1		(K) LSEGVTISYK
9	3.5e6	0450	2	0.3	1107.3	2.58	0.15	215	76	12/18	owl|p53712|itb1	+2	(R) SGEPQTFTLK
7	1.5e6	0435	2	0.3	1324.4	2.36	0.13	461	1	14/22	owl|p12607|itb0	+8	(K) TVM*PYISTTPAK
21	9.2e5	0540	2	0.3	1308.3	2.24	0.05	364	4	14/22	owl|p12607|itb0	+8	(K) TVMPYISTTPAK
27	3.1e6	0627	1	0.1	983.4	2.16	0.14	280	26	9/16	owl|p18563|itb6	+2	(R) LGFGSFVEK
13	1.5e6	0479	1	0.1	1096.5	1.99	0.14	255	10	9/18	owl|p05556|itb1		(K) LSEGVTISYK
2	8.3e5	0393	2	-1.7	1337.2	1.85	0.00	370	1	14/24	owl|p05556|itb1		(R) SNGLIC*GGNGVC*K
24	2.3e6	0572	1	0.3	1091.3	1.80	0.04	279	33	9/18	owl|p53712|itb1	+2	(K) GEVFNELVGK

F owl|U36884|HIV1U36884 16 2|5.5e5|1% (0,2,0,0,0,0) (x, 19 32, x, x, x, x)
*HIV1U36884 NID: g1079641 - Human immunodeficiency virus type 1.

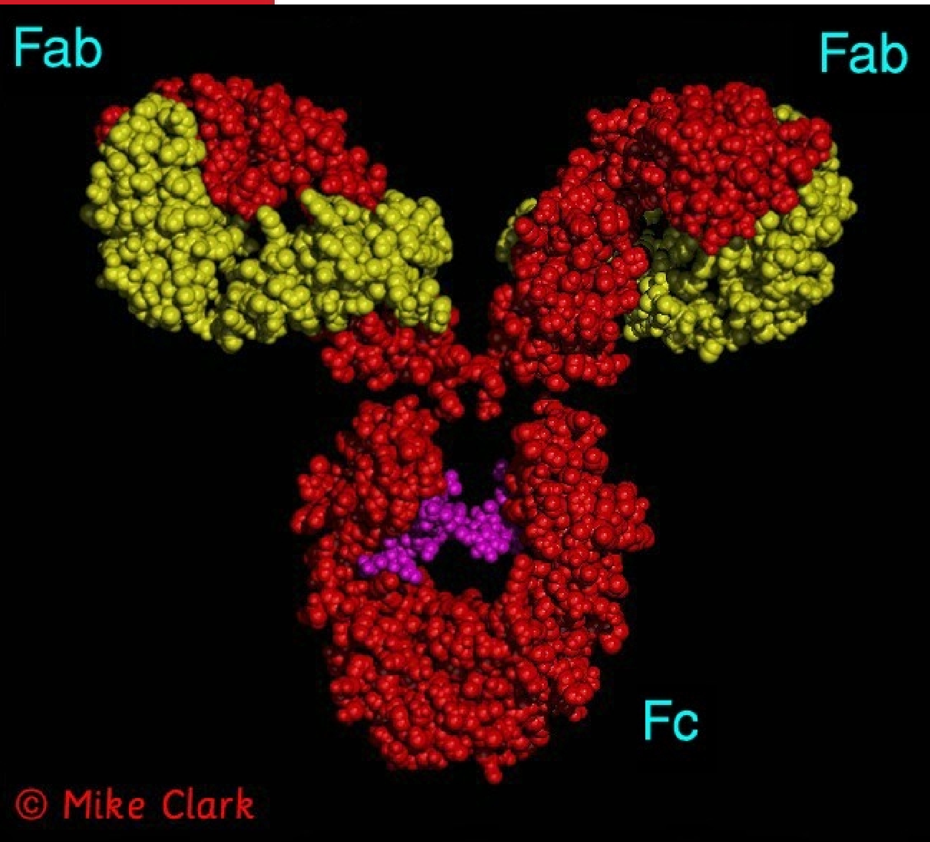
#	TIC	File	z	dM	MH+	Xcorr	dCn	Sp	RSp	Ions	Ref	()	Sequence
32	3.3e5	0699	3	-0.2	2195.5	2.44	0.04	271	93	21/68	owl|ab018306|ab		(R) EIFLFNDLLVVTKIFQKK
19	2.2e5	0534	2	0.4	1629.5	1.58	0.03	115	118	9/26	owl|u76714|rru7		(R) LADMNATIRRIDQL

G owl|AF026492|AF026492 14 2|6.3e5|1% (0,1,1,0,0,0) (x, 29, 5, x, x, x)
*AF026492 NID: g2564720 - yellow fever mosquito.

Done Internet

15 "hits" from the same protein

A real bio-LC/MS case study: anomalous heterogeneity found in early antibody production

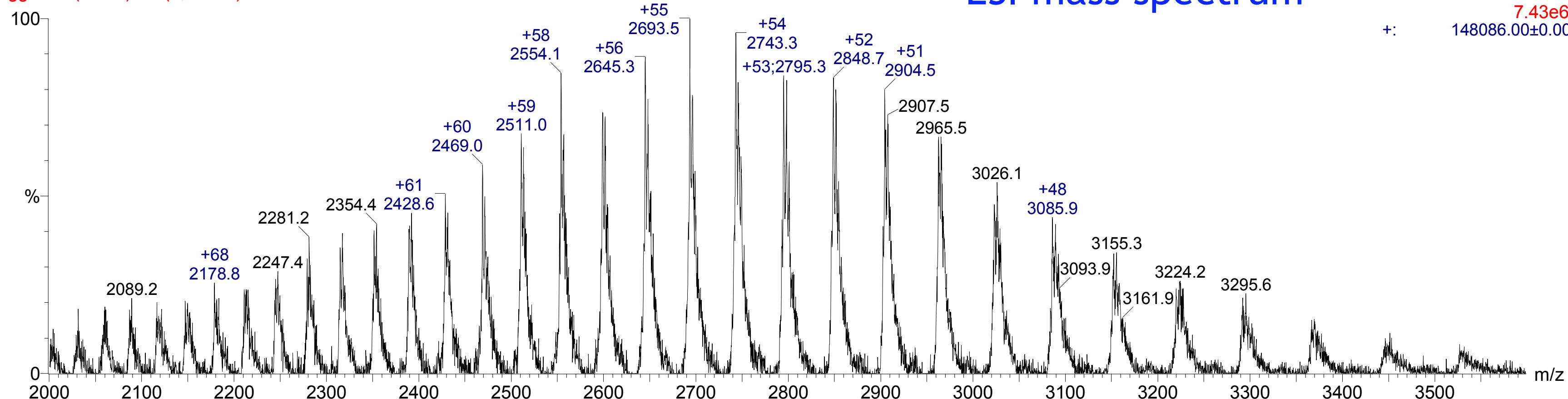


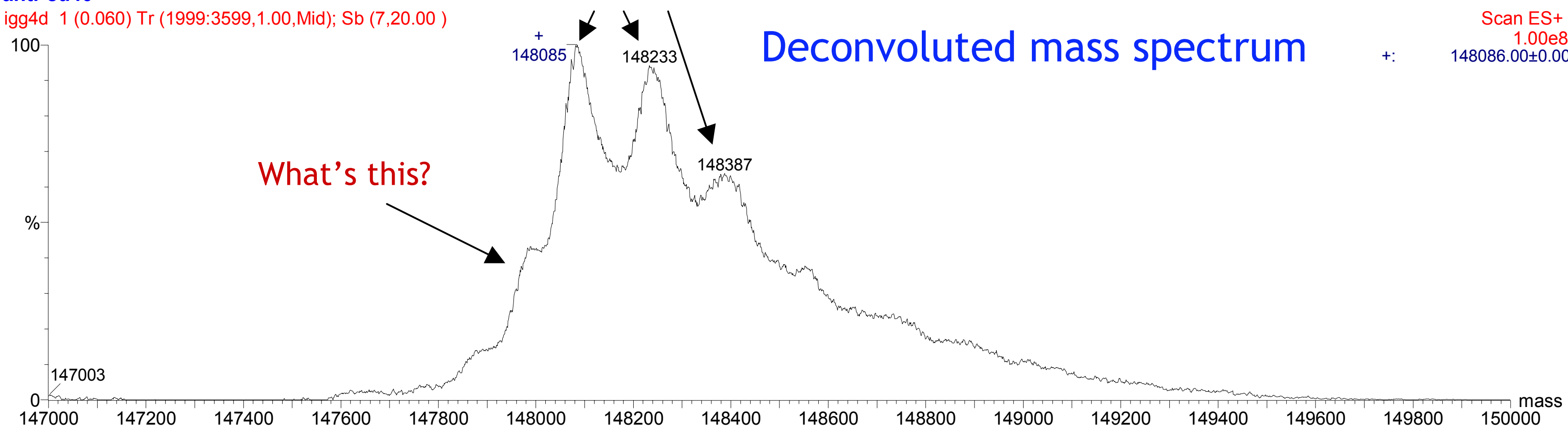
Light chain
Heavy chain
carbohydrate

- Backdrop
  - Antibody (148 kDa) being considered for therapeutic use
  - Early cell line growth and "master well" selection
  - Early development, pre-IND phase
- What happened?
  - Routine MW profiling of the intact protein by ESI/MS revealed a potential purity issue
  - Used a combination of intact protein MS (triple quad) and peptide mapping LC/MS/MS (ion trap) to solve the problem
  - Confirmed by consulting biology collaborators and considering biologically feasible options

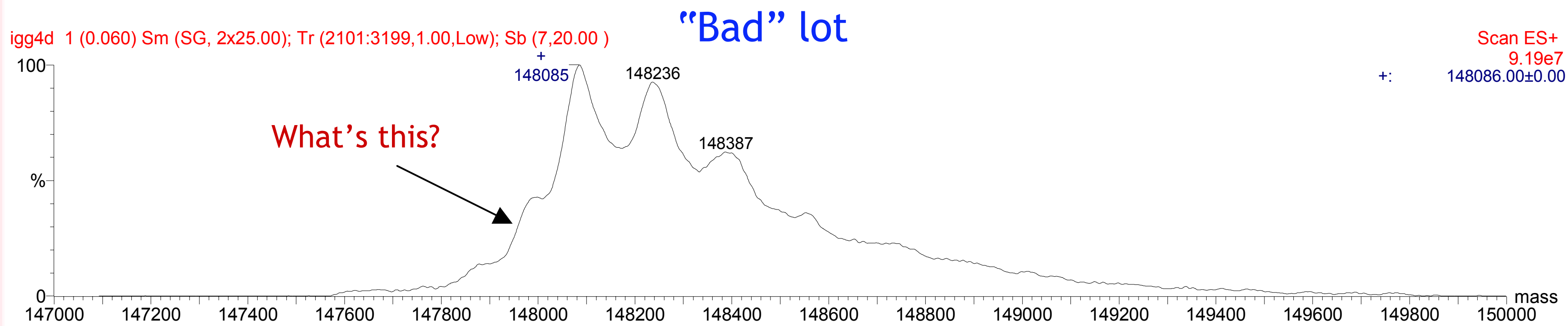
Analysis of intact antibody via ESI/MS

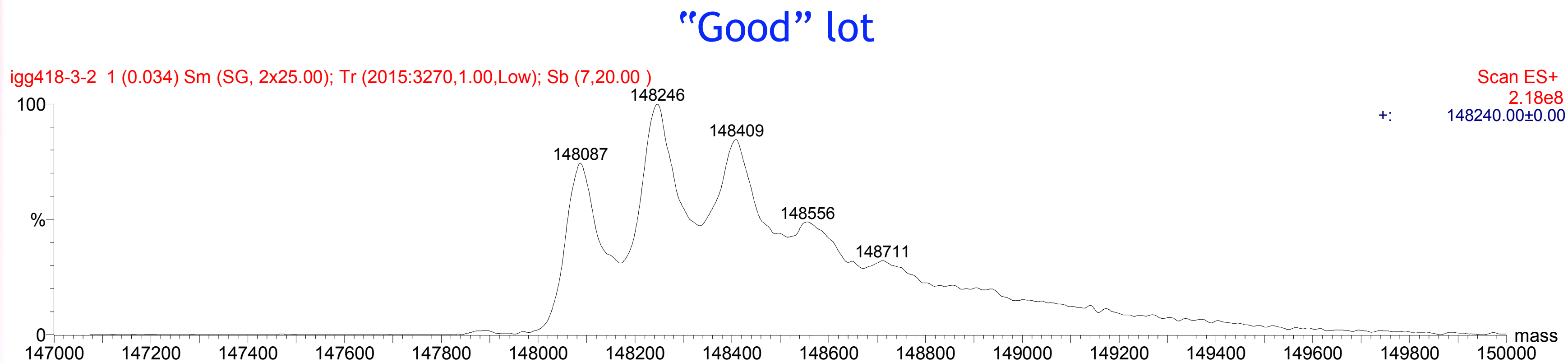
TSQ 7000 w/API-2





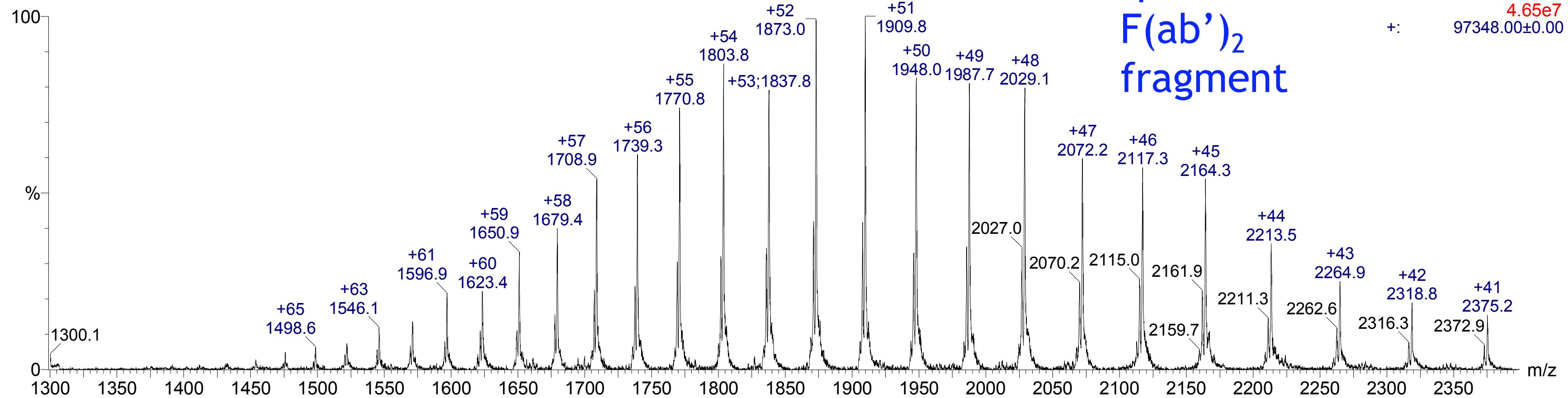
Investigation of IgG micro-heterogeneity by ESI/MS

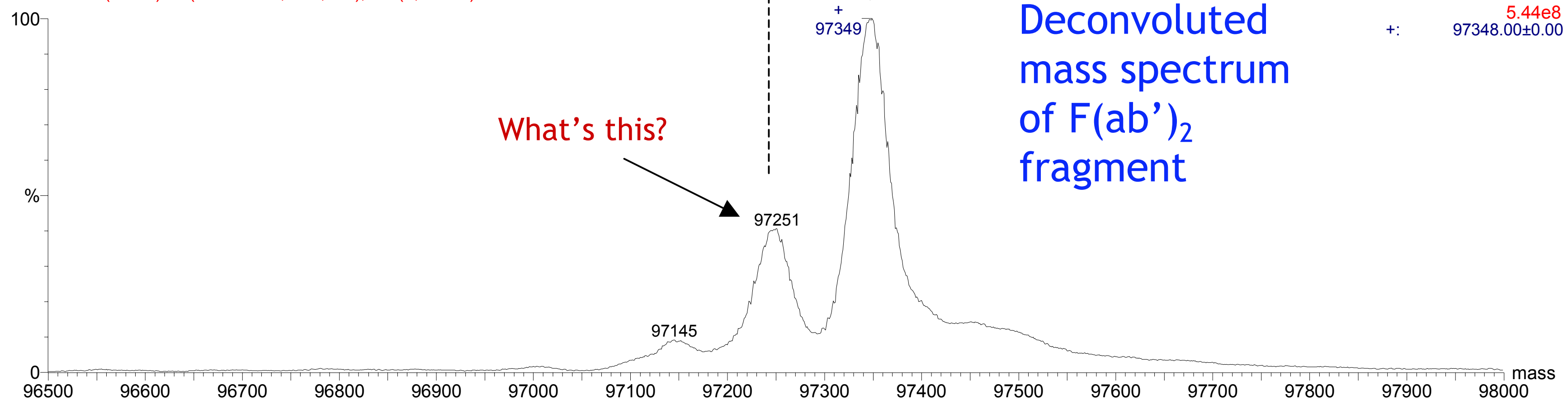




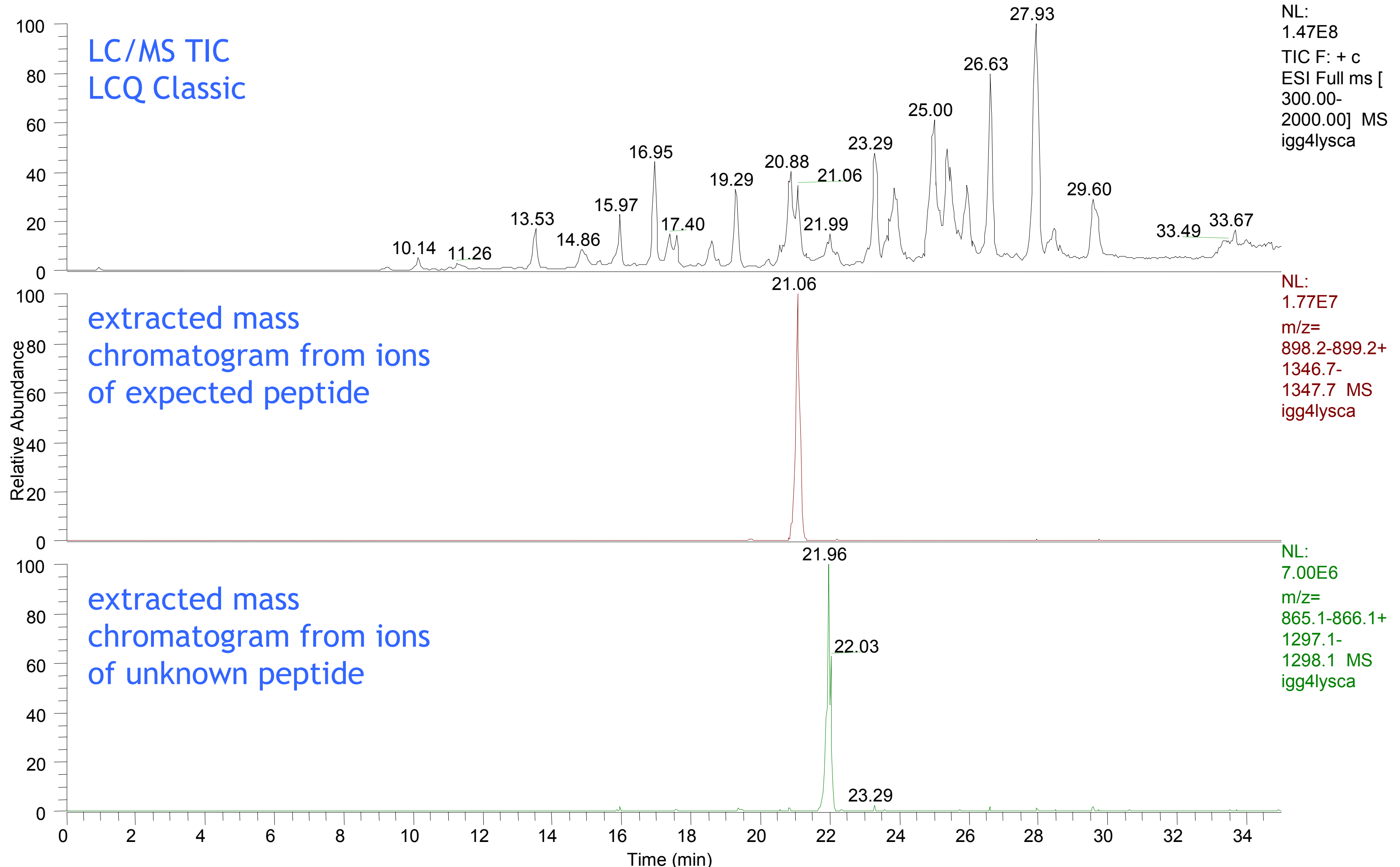
Heterogeneity follows $F(ab')_2$ fragment

mass difference appears to be ~100 Da

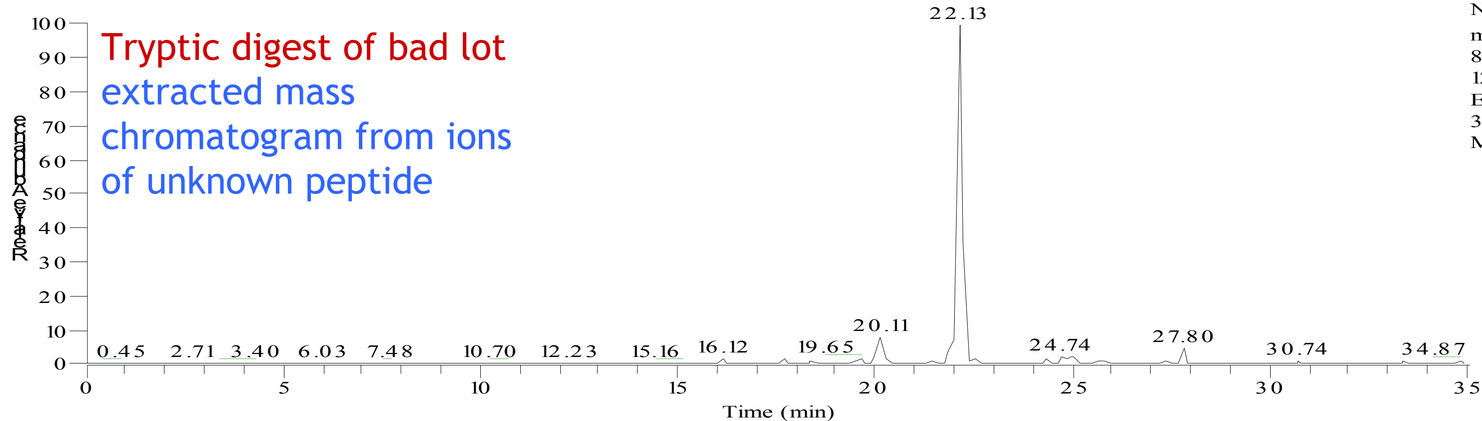


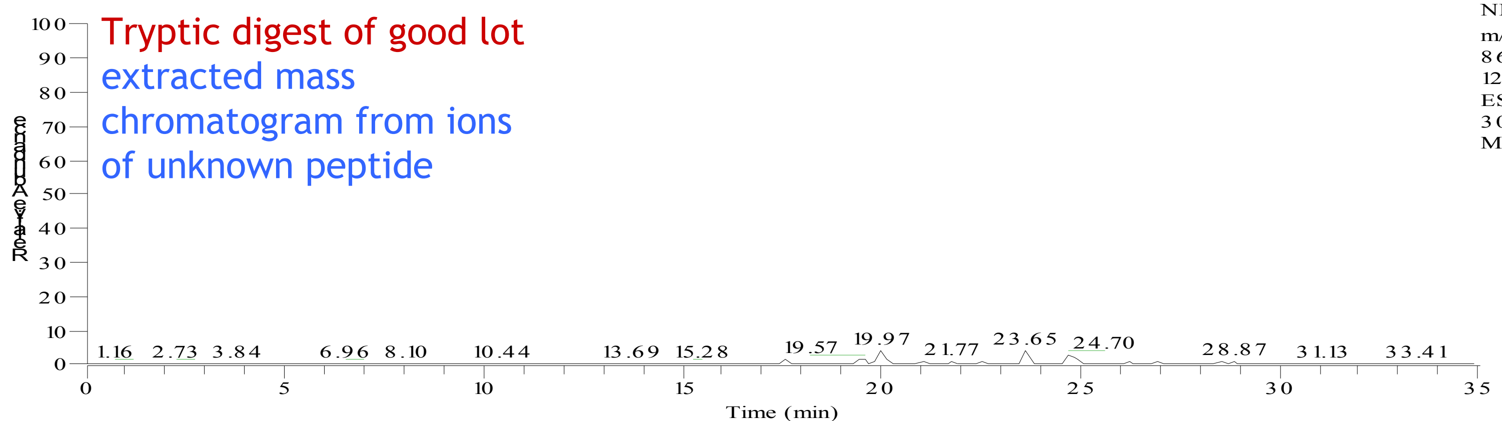


LC/MS/MS peptide mapping of LysC digest of IgG reveals anomaly in "bad" sample lot



Unknown peptide is not present in "good" lots

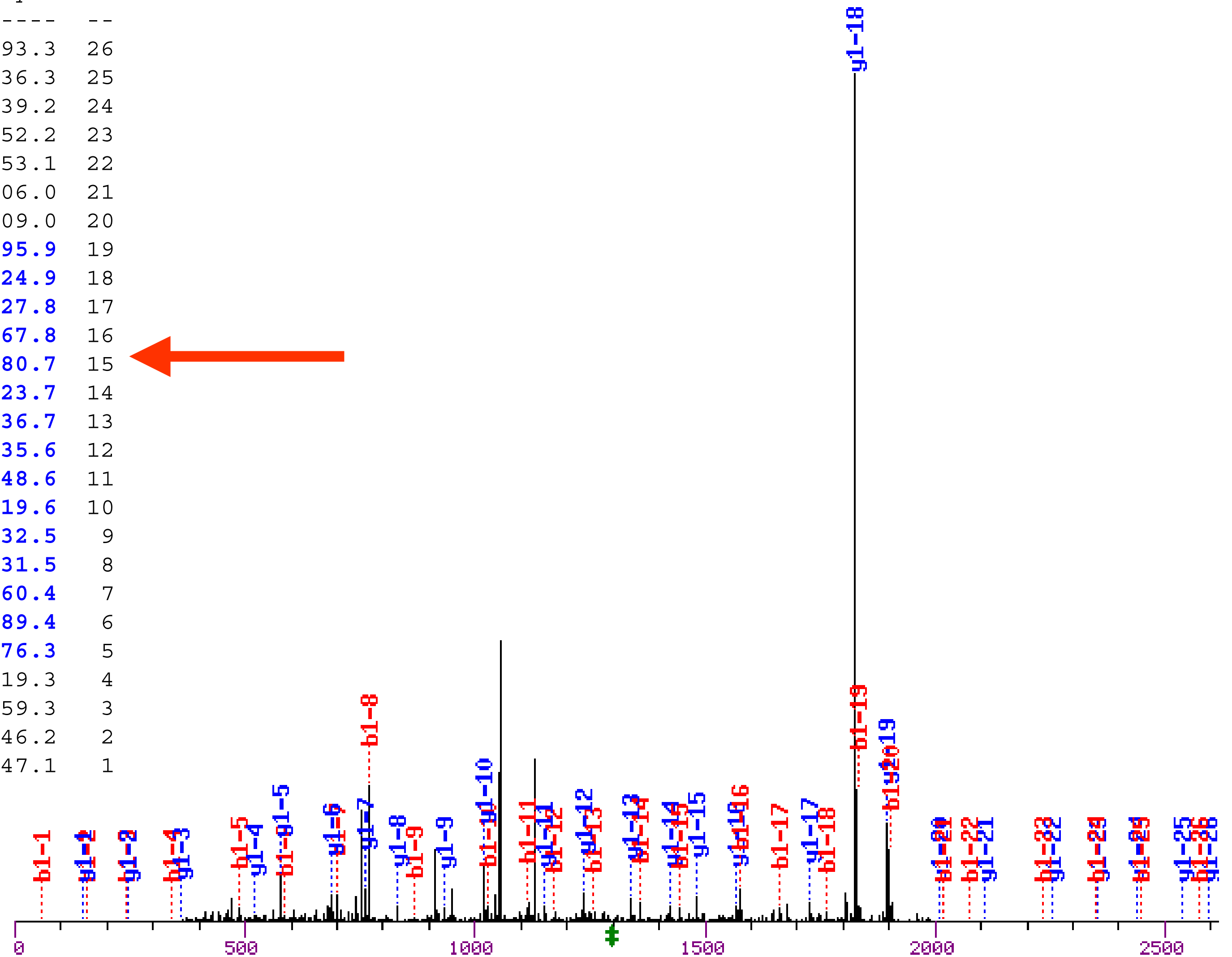




We used this simple mass chromatogram approach to screen several lots for the unknown peptide. We were able to isolate the "problem" to a single master well.

MS/MS spectrum confirming identity of unknown peptide

Mutation at position 138 on heavy chain confirmed to be Arg->Gly



Seq	#	b	y	#
G	1	58.0	2593.3	26
P	2	155.1	2536.3	25
S	3	242.1	2439.2	24
V	4	341.2	2352.2	23
F	5	488.3	2253.1	22
P	6	585.3	2106.0	21
L	7	698.4	2009.0	20
A	8	769.4	1895.9	19
P	9	866.5	1824.9	18
C	10	1026.5	1727.8	17
S	11	1113.5	1567.8	16
G	12	1170.6	1480.7	15
S	13	1257.6	1423.7	14
T	14	1358.6	1336.7	13
S	15	1445.7	1235.6	12
E	16	1574.7	1148.6	11
S	17	1661.8	1019.6	10
T	18	1762.8	932.5	9
A	19	1833.8	831.5	8
A	20	1904.9	760.4	7
L	21	2018.0	689.4	6
G	22	2075.0	576.3	5
C	23	2235.0	519.3	4
L	24	2348.1	359.3	3
V	25	2447.2	246.2	2
K	26	2575.3	147.1	1

How did we figure it out?

- We collected MS of the intact antibody and saw some "odd" looking peaks, lower by ~100 Da.
- We looked at the $F(ab')_2$ fragment, the odd looking peaks were present, therefore modification must be on this portion of the molecule – I.e., not on the Fc portion on the heavy chain
- Reduced/alkylated, digested with LysC (and trypsin)
- Performed data-dependent LC/MS/MS expts. on ion trap
- Used SEQUEST to identify all of the known peptides
- Looked for signals that did not correlate
- Found a peptide 99 Da lower in mass which eluted near a known peptide, inspected MS/MS spectrum
- Considered AA modifications that would give this result
- Arg -> Gly substitution made the most sense
  - MS/MS spectrum verification
  - Peptide is not cleaved in the tryptic digest, suggests that Arg has been converted to something else
  - From the genetic code, modification of a single base in the codon could cause this to happen (later isolated to a single master well)

Only 2 possibilities for –99 Da mass shift

http://prowl.rockefeller.edu/aainfo/mutation.html

Mutation Mass Shifts - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Real.com

Address http://prowl.rockefeller.edu/aainfo/mutation.html Go Links

Mutation Mass Shifts

Masses changes between various amino acids)

1. Residues DOWN the left indicate the EXPECTED residues.
2. Residues ACROSS the top indicate the MUTANT residues.

		Gly	Ala	Ser	Pro	Val	Thr	Cys	Leu/ Ile	Asn	Asp	Gln/ Lys	Glu	Met	His	Phe	Arg	Tyr	Trp
		57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186
Gly	57		14	30	40	42	44	46	56	57	58	71	72	74	80	90	99	106	129
Ala	71	-14		16	26	28	30	32	42	43	44	57	58	60	66	76	85	92	115
Ser	87	-30	-16		10	12	14	16	26	27	28	41	42	44	50	60	69	76	99
Pro	97	-40	-26	-10		2	4	6	16	17	18	31	32	34	40	50	59	66	89
Val	99	-42	-28	-12	-2		2	4	14	15	16	29	30	32	38	48	57	64	87
Thr	101	-44	-30	-14	-4	-2		2	12	13	14	27	28	30	36	46	55	62	85
Cys	103	-46	-32	-16	-6	-4	-2		10	11	12	25	26	28	34	44	53	60	83
Leu/Ile	113	-56	-42	-26	-16	-14	-12	-10		1	2	15	16	18	24	34	43	50	73
Asn	114	-57	-43	-27	-17	-15	-13	-11	-1		1	14	15	17	23	33	42	49	72
Asp	115	-58	-44	-28	-18	-16	-14	-12	-2	-1		13	14	16	22	32	41	48	71
Gln/Lys	128	-71	-57	-41	-31	-29	-27	-25	-15	-14	-13		1	3	9	19	28	35	58
Glu	129	-72	-58	-42	-32	-30	-28	-26	-16	-15	-14	-1		2	8	18	27	34	57
Met	131	-74	-60	-44	-34	-32	-30	-28	-18	-17	-16	-3	-2		6	16	25	32	55
His	137	-80	-66	-50	-40	-38	-36	-34	-24	-23	-22	-9	-8	-6		10	19	26	49
Phe	147	-90	-76	-60	-50	-48	-46	-44	-34	-33	-32	-19	-18	-16	-10		9	16	39
Arg	156	-99	-85	-69	-59	-57	-55	-53	-43	-42	-41	-28	-27	-25	-19	-9		7	30
Tyr	163	-106	-92	-76	-66	-64	-62	-60	-50	-49	-48	-35	-34	-32	-26	-16	-7		23
Trp	186	-129	-115	-99	-89	-87	-85	-83	-73	-72	-71	-58	-57	-55	-49	-39	-30	-23	

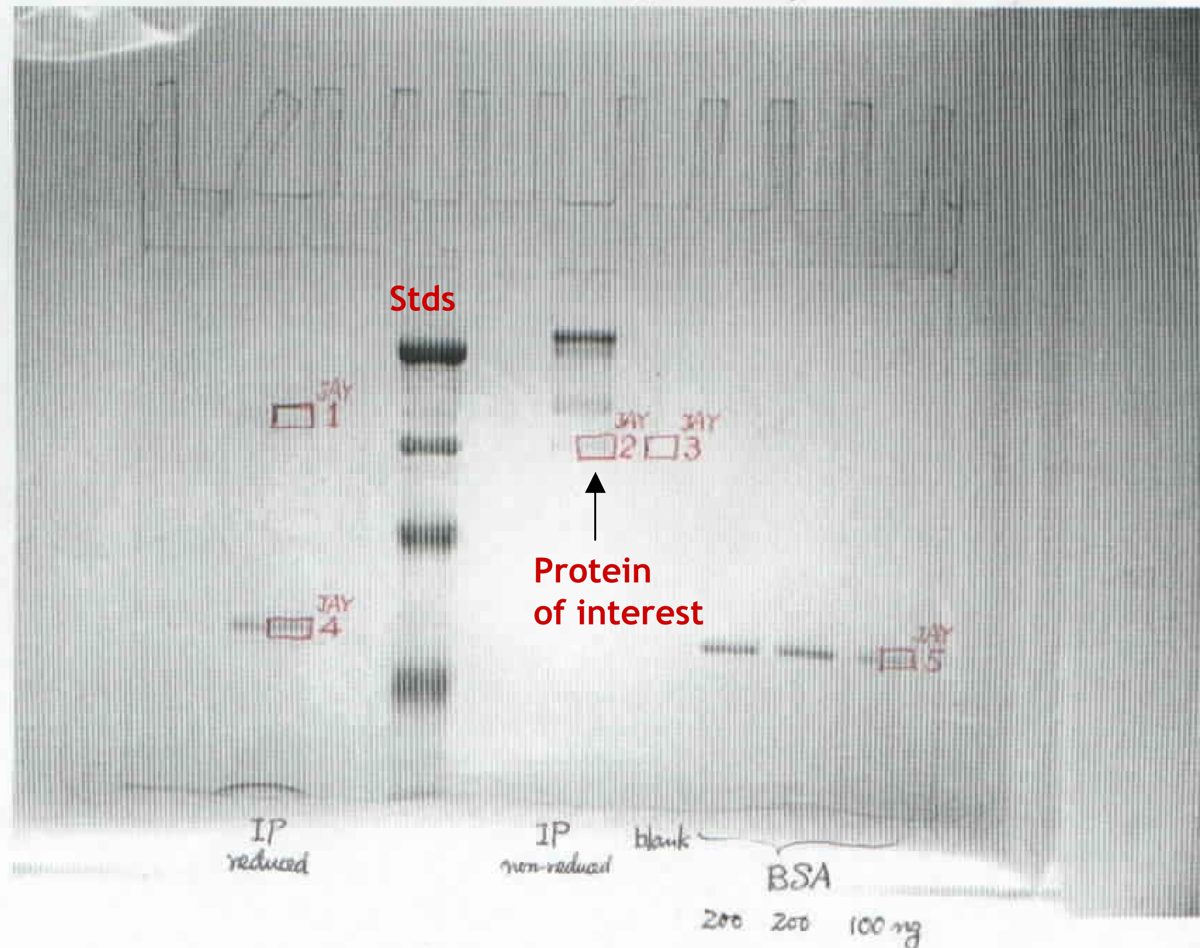
kclauser@rafael.ucsf.edu 5/1/95

Done Internet

Arg -> Gly
Trp -> Ser

A real bio-LC/MS case study: identification of immunoprecipitated protein separated by SDS-PAGE

Photo of 1-D SDS-PAGE of immunoprecipitate from human fibrosarcoma cell lysate
~200 fmoles TOTAL loaded

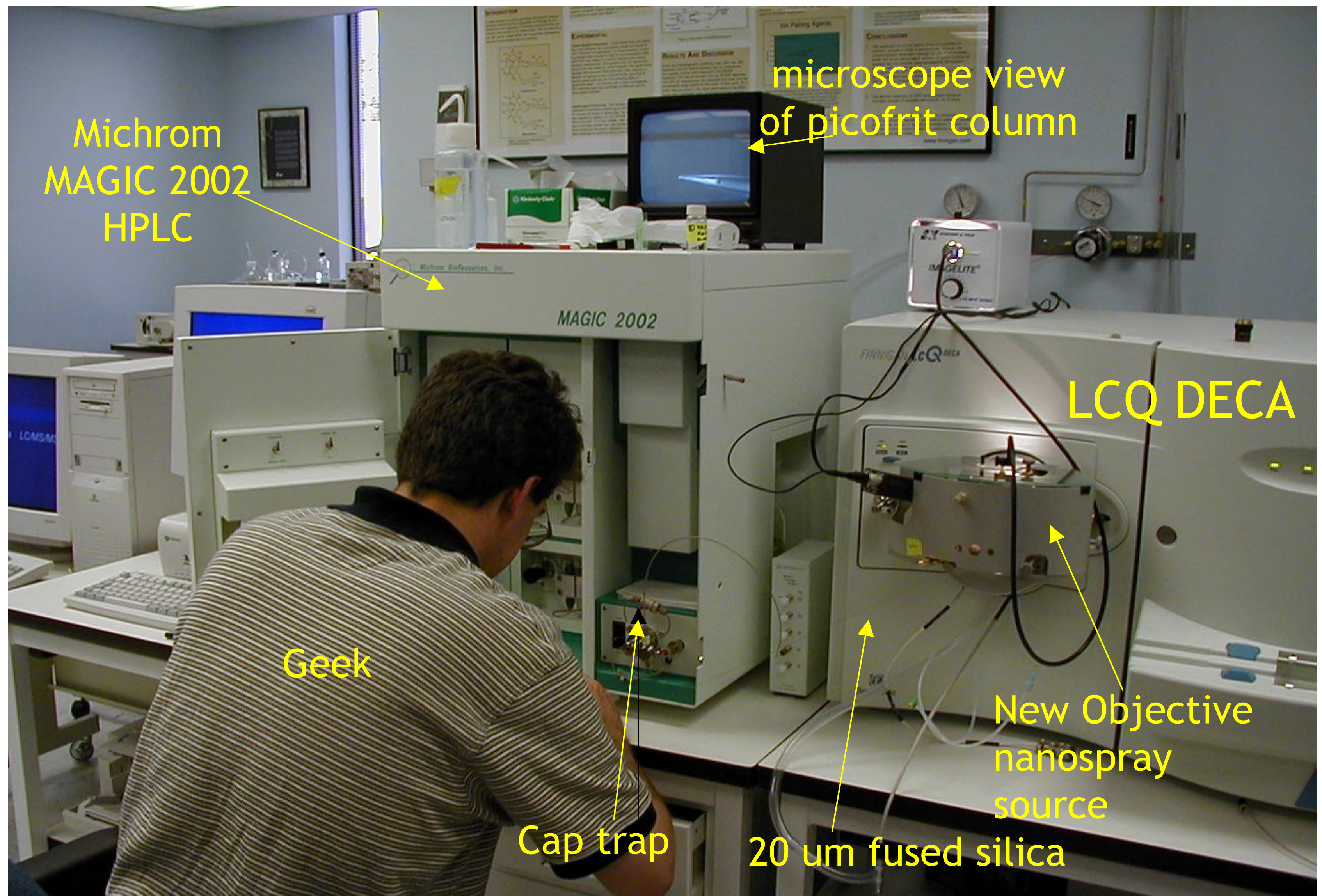


Protein i.d. case study

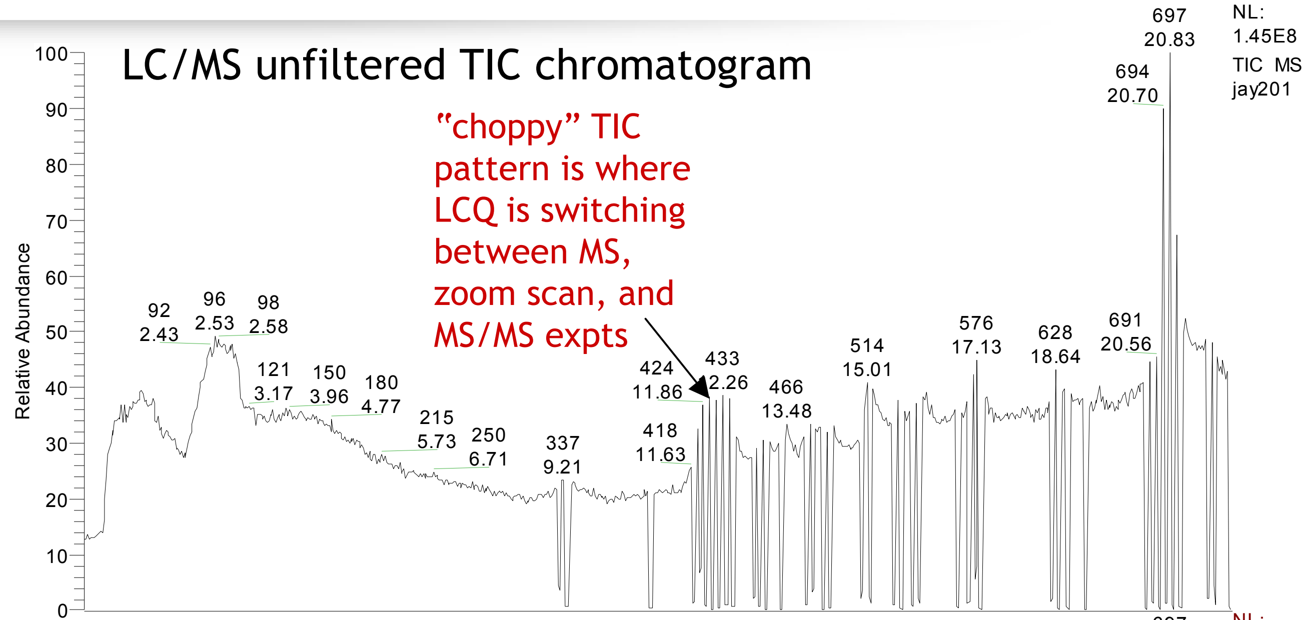
Experimental Approach

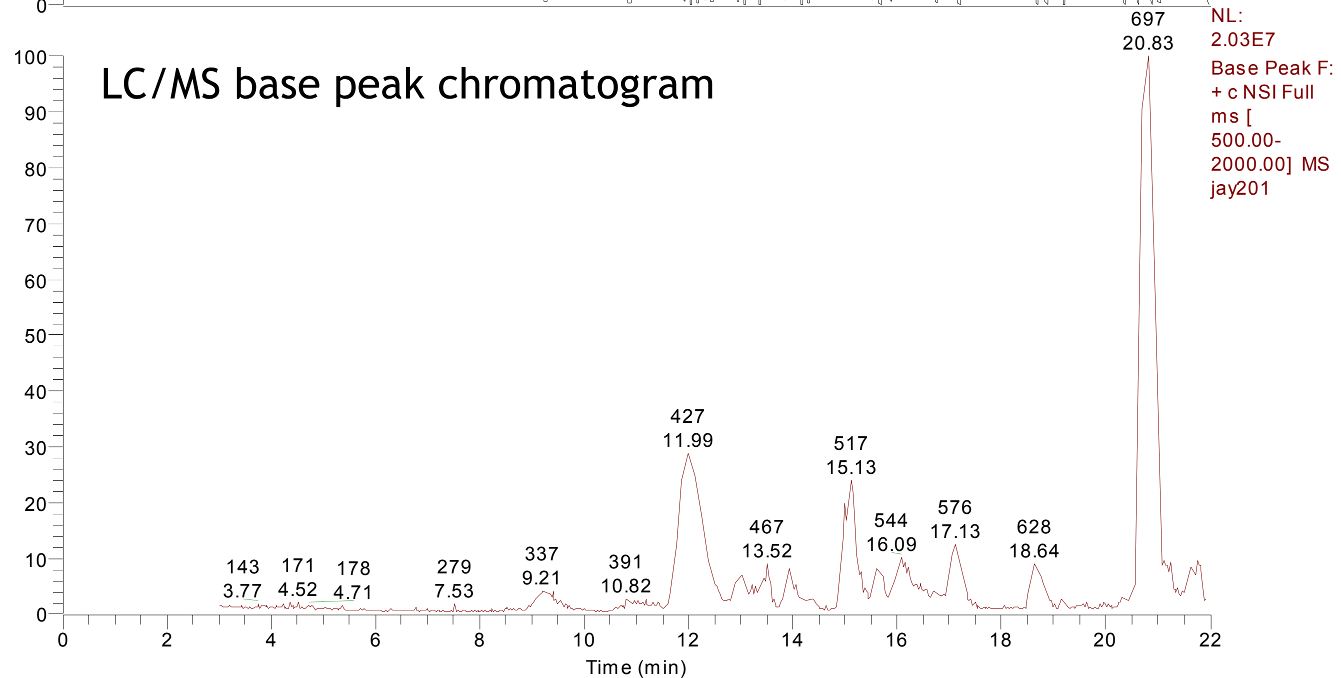
- Excise protein from gel
- Perform in-gel digest
- Inject onto nano-LC (use nano-scale trapping column to preconcentrate sample) on-line with ion trap MS
- Perform data-dependent "triple-play" scanning: MS scan, hi-res zoom scan, MS/MS scan
- Analyze results using TurboSEQUEST protein database cross-correlation program

Photo of Experimental nano-LC/MS/MS system

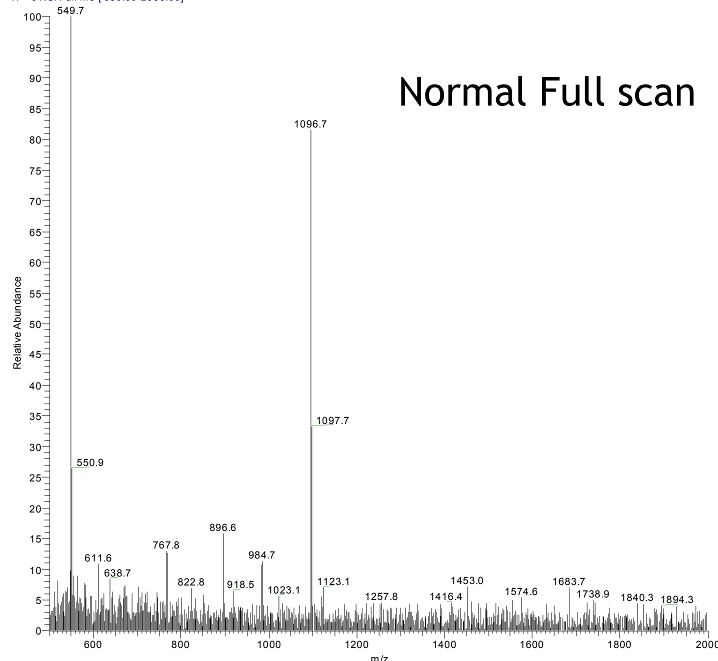


Data-dependent LC/MS/MS analysis of in-gel digest sample

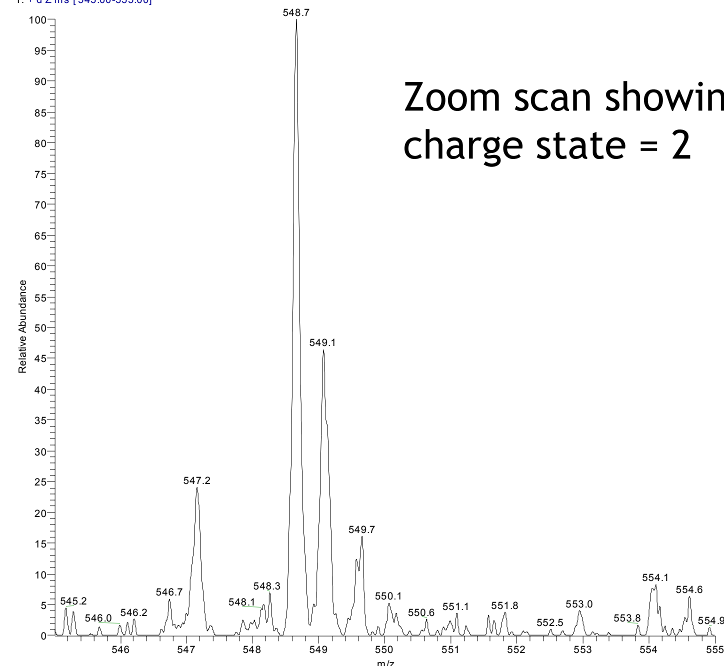




Example of triple-play data from in-gel digest sample

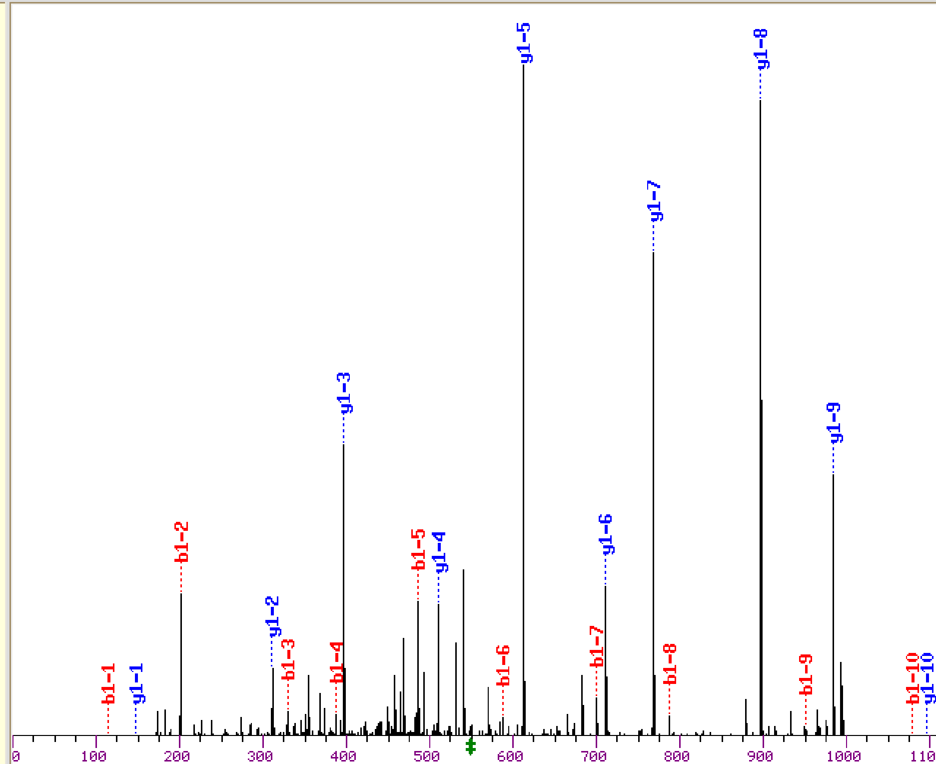


Normal Full scan



Zoom scan showing charge state = 2

Seq	#	b	y	(+1)
L	1	114.1	1096.6	10
S	2	201.1	983.5	9
E	3	330.2	896.5	8
G	4	387.2	767.4	7
V	5	486.3	710.4	6
T	6	587.3	611.3	5
I	7	700.4	510.3	4
S	8	787.4	397.2	3
Y	9	950.5	310.2	2
K	10	1078.6	147.1	1



MS/MS spectrum of m/z 549

SEQUEST summary from LC/MS/MS analysis of in-gel digest sample

Sequest Summary - Microsoft Internet Explorer

Address: http://localhost/cgi-bin/runsummary.pl?directory=mhailjay2_itb1&sort=consensus

Sequest Summary

Setup CreateDTA VuDTA RunSequest Status Summary Utilities Home

Sample: Hail, M. (JAY2_ITB1) JAY2 meh   Db: nr (05/12/2000)   Inspector View Info   Mass: Mono
Datafiles: jay201 (06/03/2000-06/03/2000)   Dir: mhailjay2_itb1   Enz: Trypsin   Tot: 13|13|3.1e7
Intensity: Full Zoom MS2   Diff Mods: 57 C 16 M X
Consensus Chron Max rank: 3 Max list: 5 Controls: Show DTA VCR

A gi|4504767|ref|NP_002202.1|| 88 13|3.1e7|100% (8,1,0,0,0,0) (1 2 3 4 5 6 7 8 9 10 11 13, 12, x, x, x, x)
*integrin, beta 1 (fibronectin receptor, betapolypeptide, antigen CD29 includes MDF2, MSK12)gi|124963|sp|P05556|ITB1_HUMAN FIBRONECTIN RECEPTOR BETA SUBUNIT PRECURSOR (INTEGRIN BETA-1) (CD29)(INTEGRIN VLA-4 BETA SUBUNIT)gi|87429|pir||B27079 fibronectin receptor beta chain precursor - humangi|31442

#	TIC	File	z	dM	MH+	Xcorr	dCn	Sp	RSp	Ions	Ref	()	Sequence
13	2.0e6	0690	2	-0.4	1267.0	3.92	0.43	1353	1	17/20	gi|4504767	+5	(K)SLGTDLMNEMR
4	8.9e5	0438	3	-1.2	1846.1	3.84	0.26	1453	1	29/60	gi|4504767		(R)DKLPQPVQPDPVSHC*K
2	8.0e5	0432	2	0.1	1536.4	3.26	0.32	1042	1	15/20	gi|4504767	+7	(K)FC*EC*DNFNC*DR
8	5.1e6	0511	2	-0.7	1223.3	3.23	0.41	1131	1	16/18	gi|4504767	+14	(K)WDTGENPIYK
7	5.5e6	0482	2	-0.5	1097.1	3.08	0.36	1202	1	16/18	gi|4504767		(K)LSEGVTISYK
9	1.5e6	0531	2	0.2	1282.4	3.03	0.37	1134	1	16/20	gi|4504767	+5	(K)SLGTDLMNEM#R
1	2.2e6	0420	2	0.2	1298.4	2.73	0.22	1201	1	16/20	gi|4504767	+5	(K)SLGTDLM#NEM#R
3	1.5e6	0435	2	0.3	1324.4	2.66	0.20	405	2	14/22	gi|4504767	+10	(K)TVM#PYISTTPAK
5	3.5e6	0450	2	0.3	1107.3	2.65	0.24	470	1	14/18	gi|4504767	+2	(R)SGEPQTFTLK
10	9.2e5	0540	2	0.3	1308.3	2.35	0.21	507	1	15/22	gi|4504767	+10	(K)TVMPYISTTPAK
11	2.3e6	0572	1	0.3	1091.3	2.09	0.17	348	11	10/18	gi|4504767	+2	(K)GEVFNELVGK
12	3.1e6	0627	1	0.1	983.4	2.04	0.12	557	1	12/16	gi|4504775	+3	(R)LGFGSFVEK
6	1.5e6	0479	1	0.1	1096.5	2.01	0.19	322	5	10/18	gi|4504767		(K)LSEGVTISYK

Pull to Top

Confirmed hits for integrin beta-1

A gi|4504767|ref|NP_002202.1|| 88 (8,1,0,0,0,0) (1 2 3 4 5 6 7 8 9 10 11 13, 12, x, x, x, x)
*integrin, beta 1 (fibronectin receptor, betapolypeptide, antigen CD29 includes MDF2, MSK12)gi|124963|sp|P05556|ITB1_HUMAN FIBRONECTIN RECEPTOR BETA SUBUNIT PRECURSOR (INTEGRIN BETA-1) (CD29)(INTEGRIN VLA-4 BETA SUBUNIT)gi|87429|pir||B27079 fibronectin receptor beta chain precursor - humangi|31442

B gi|1708573|sp|P53712|ITB1_BOVIN 68 (6,1,0,0,0,0) (1 2 3 5 8 9 10 11 13, 12, x, x, x, x)

C gi|1708574|sp|P53713|ITB1_FELCA 58 (5,1,0,0,0,0) (2 3 5 8 10 11, 12, x, x, x, x)

D gi|124964|sp|P09055|ITB1_MOUSE 48 (4,1,0,0,0,0) (1 2 3 8 9 10 13, 12, x, x, x, x)
gi|1352494, gi|72070, gi|762977

E gi|124961|sp|P12607|ITB0_XENLA 28 (2,1,0,0,0,0) (3 8 10, 12, x, x, x, x)
gi|124962, gi|124965

*Database nr.fasta unavailable for sequence descriptions.

http://www.ncbi.nlm.nih.gov/blast/blast.cgi?SEQUENCE=KFCECDNFNCDR&PROGRAM=blastp&DATALIB=nr&EXPECT=1000&ALI(Local intranet

Sequence coverage map of integrin beta-1 from LC/MS/MS analysis

Flicka - Microsoft Internet Explorer

File

Address http://localhost/cgi-bin/flicka.pl?Db=d:/Xcalibur/database/nr.fasta&MassType=1&Pep=SLGTDLMNEMR+GEVFNELVGK+TVMPYISTTPAK+FCECDNFNCDR+ Go Links

Flicka

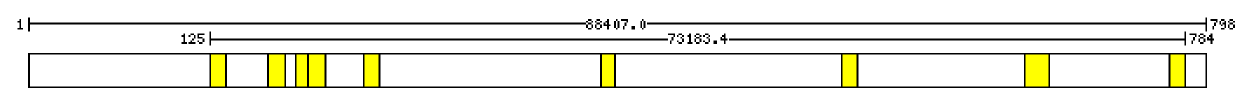
Setup CreateDTA VuDTA RunSequest Status Summary Utilities Home

Send to: PEPCUT PEPSTAT BLAST **NCBI:** SEQUENCE ABSTRACT

>**gi|4504767**|ref|NP_002202.1|| integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)□gi|124963|sp|P05556|ITB1_HUMAN FIBRONECTIN RECEPTOR BETA SUBUNIT PRECURSOR (INTEGRIN BETA-1) (CD29) (INTEGRIN VLA-4 BETA SUBUNIT)□gi|87429|pir||B27079 fibronectin receptor beta chain precursor - human□gi|31442|emb|CAA30790.1| (X07979) integrin beta 1 subunit precursor [Homo sapiens]

MNLQPIFWIG LISSVCCVFA QTDENRCLKA NAKSCGECIQ AGPNCGWCTN STFLQEGMPT SARCDDLEAL KKKGCPPDDI
ENPRGSKDIK KNKNVTNRSK GTAEKLKPED IHQIQPQQLV LRLR**SGEPQT FTLK**FKRAED YPIDLYYLMD LSYSMKDDLE
NVK**SLGTDLM NEMR**RITSDF R**IGFGSFVEK TVMPYISTTP AK**LRNPCTSE QNCTTPFSYK NVLSLTNK**GE VFNELVGK**QR
ISGNLDSPEG GFDAIMQVAV CGSLIGWRNV TRLLVFSTDA GFHFAGDGKL GGIVLPNDGQ CHLENNMYTM SHYYDYPSIA
HLVQKLSENN IQTIFAVTEE FQPVYKELKN LIPKSAVGTL SANSSNVIQL IIDAYNSLSS EVILENGK**LS EGVTISYK**SY
CKNGVNGTGE NGRKCSNISI GDEVQFEISI TSNKCPKKDS DSFKIRPLGF TEEVEVILQY ICECECQSEG IPESPKCHEG
NGTFECGACR CNEGRVGRHC ECSTDEVNSE DMDAYCRKEN SSEICSNNGE CVCGQCVCRK RDNTNEIYSG K**FCECDNFNC**
DRSNGLICGG NGVCKCRVCE CNPNYTGSAC DCSLDTSTCE ASNGQICNGR GICECGVCKC TDPKFQGQTC EMCQTCLGVC
AEHKECVQCR AFNKGEKKDT CTQECSYFNI TKVESR**DKLP QPVQPDPVSH CK**EKDVDDCW FYFTYSVNGN NEVMVHVVEN
PECPTGPDII PIVAGVVAGI VLIGLALLLI WKLLMIIHDR REFAKFEKEK MNAK**WDTGEN PIYK**SAVTTV VNPKYEGK

Mass (mono): 88407.0 **Identifier:** gi|4504767 **Database:** d:/Xcalibur/database/nr.fasta
Protein Coverage: 99/798 = **12.4%** by amino acid count, 11077.3/88407.0 = **12.5%** by mass



Sort by: Sequence ☐ Position ☐

	Sequence	Position
PEPSTAT	DKLPQPVQPDPVSHCK	677 - 692
PEPSTAT	FCECDNFNCDR	552 - 562
PEPSTAT	GEVFNELVGK	229 - 238
PEPSTAT	IGFGSFVEK	182 - 190
PEPSTAT	LSEGVTISYK	389 - 398
PEPSTAT	SGEPQTFTLK	125 - 134
PEPSTAT	SLGTDLMNEMR	164 - 174
PEPSTAT	TVMPYISTTPAK	191 - 202
PEPSTAT	WDTGENPIYK	775 - 784

Done Local intranet

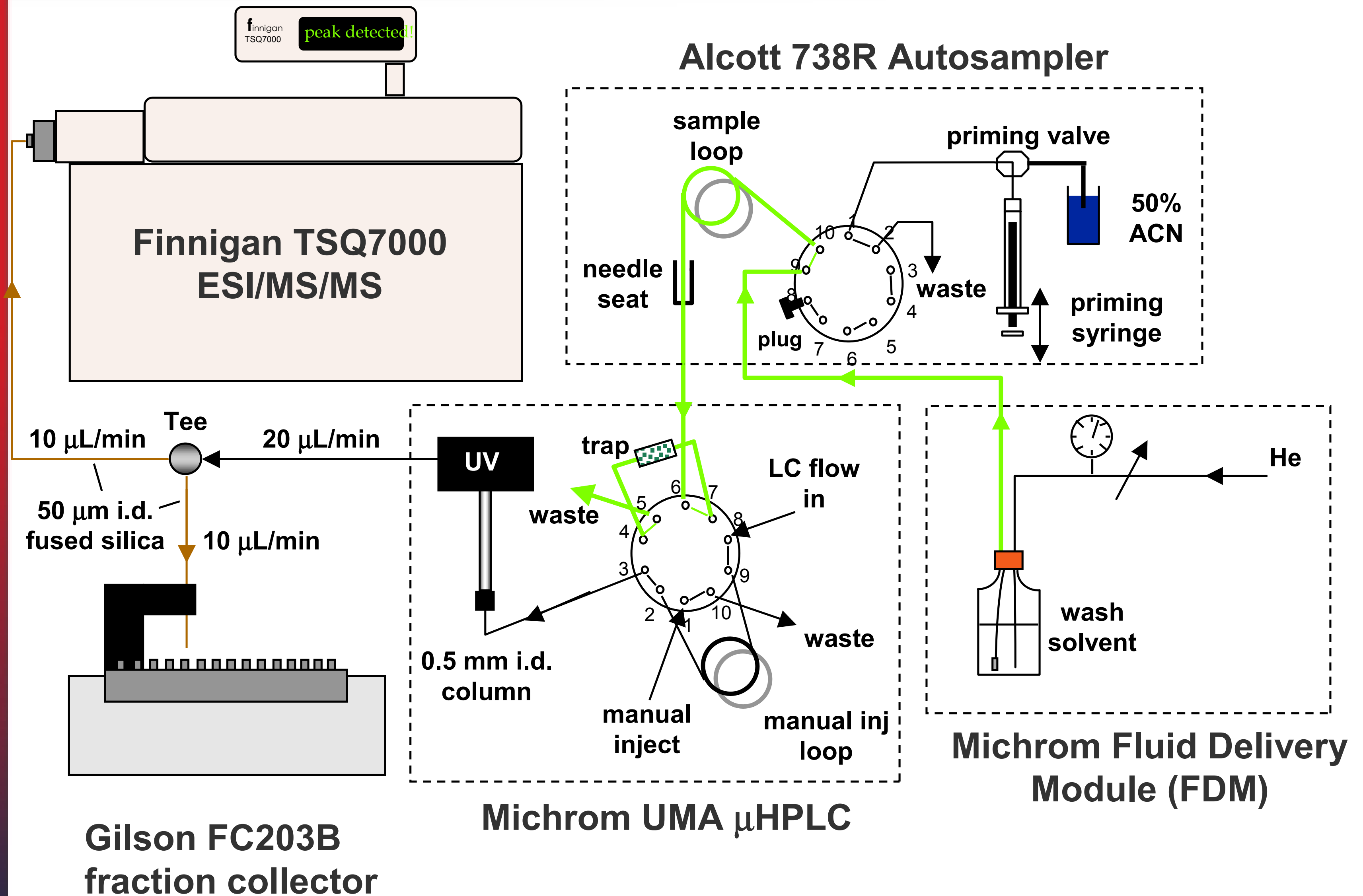
Protein i.d. case study conclusions

- Protein confirmed as integrin beta-1
- All experiments were performed in a single nano-LC/MS/MS run using data-dependent scanning
- 9 different peptides from the protein were confirmed as database search hits
- Several oxidized Met peptides were observed (likely the result of the in-gel digest procedure)

Opportunities for Automation

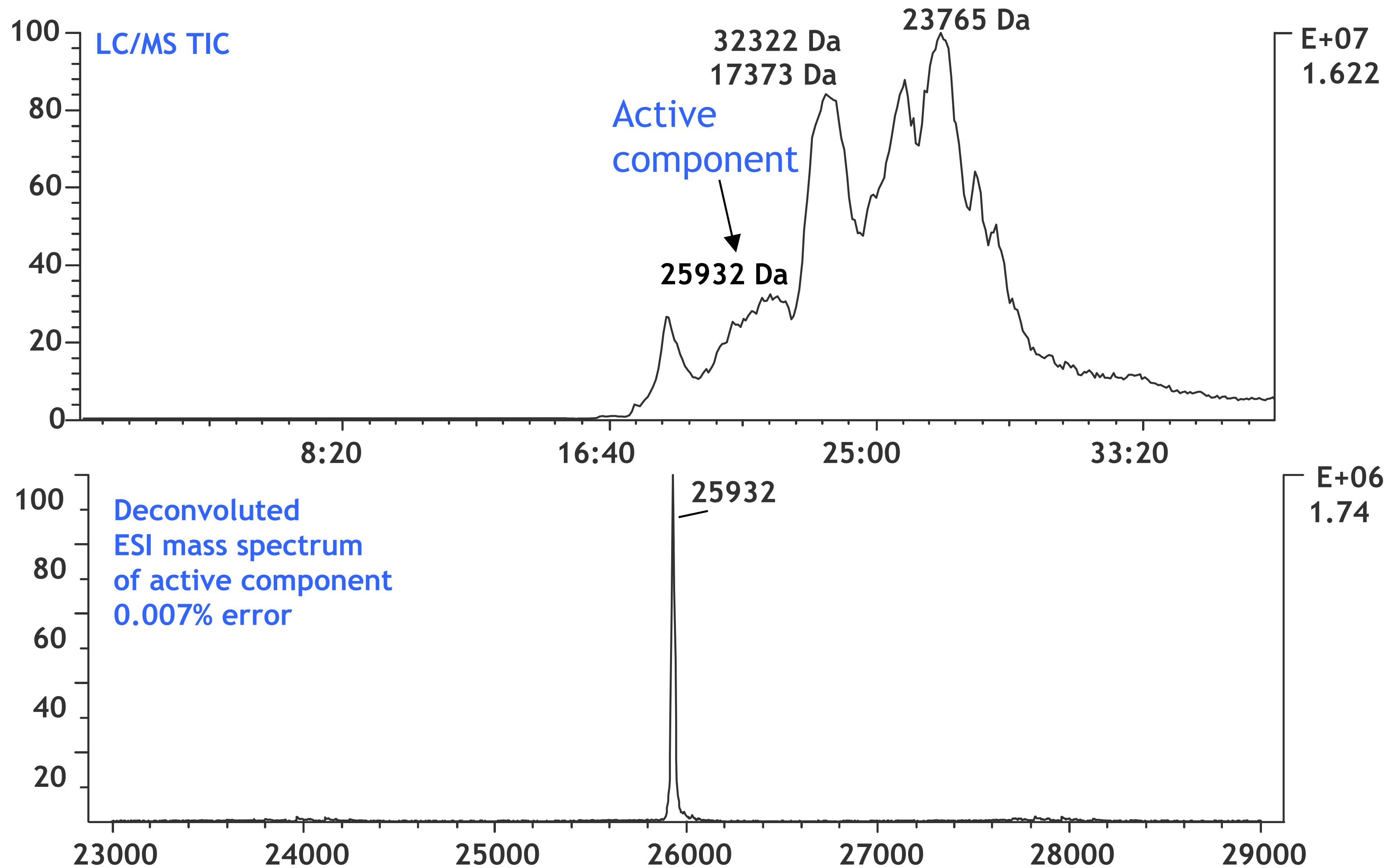
- Automated sample handling
  - Minimize sample losses through sample transfer and handling steps
  - Concentrate, desalt, remove contaminants on-line
  - Digest on-line (combine with data-dependent triple play for completely automated peptide mapping system)
  - Rank large-small molecule affinities on-line
- Data-manipulation (bioinformatics – a whole new area)
  - What to do with SEQUEST output
  - Automated in-tact protein analysis

Schematic of Automated Sample Cleanup and Fraction Collection System

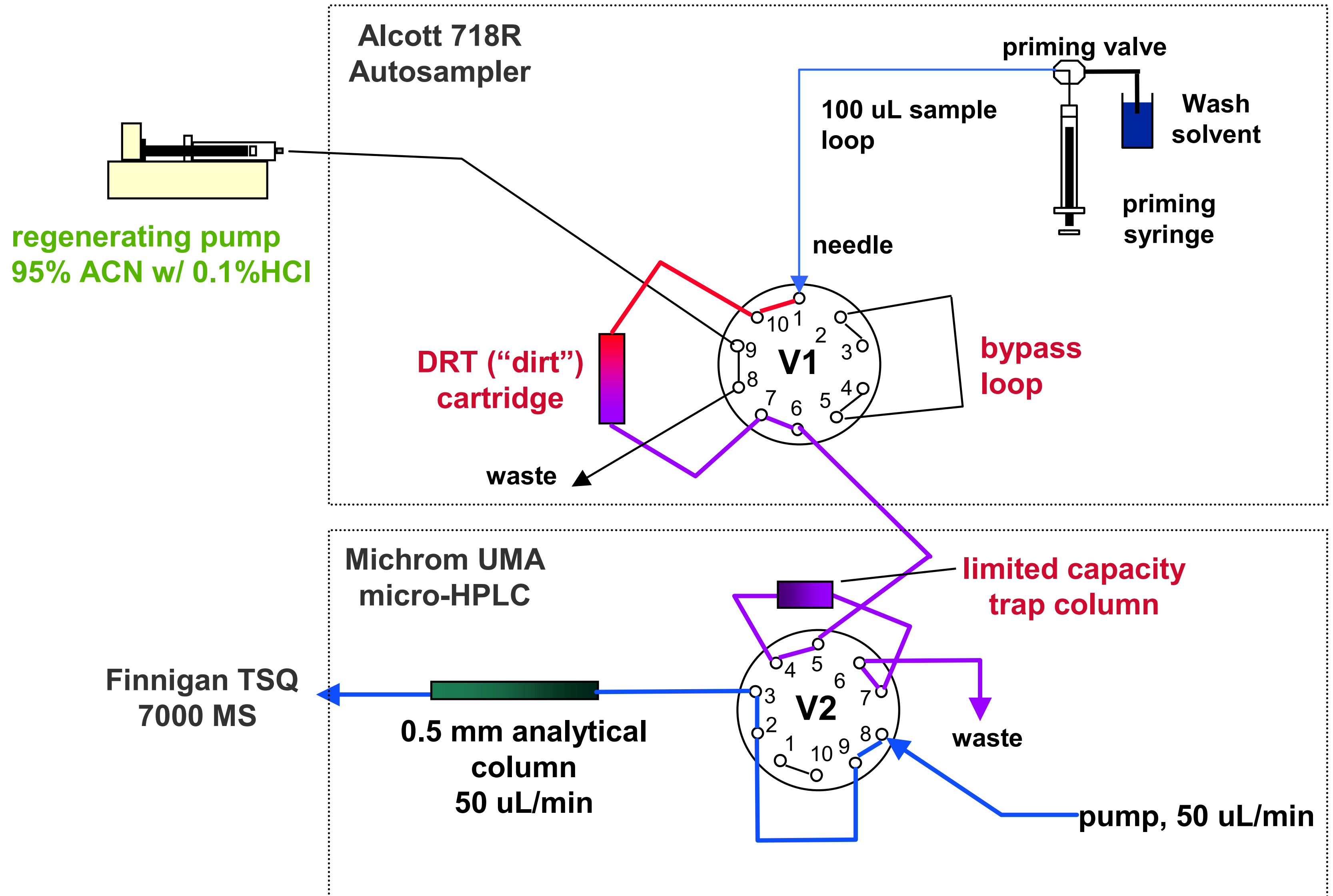


Automated protein sample preconcentration/cleanup

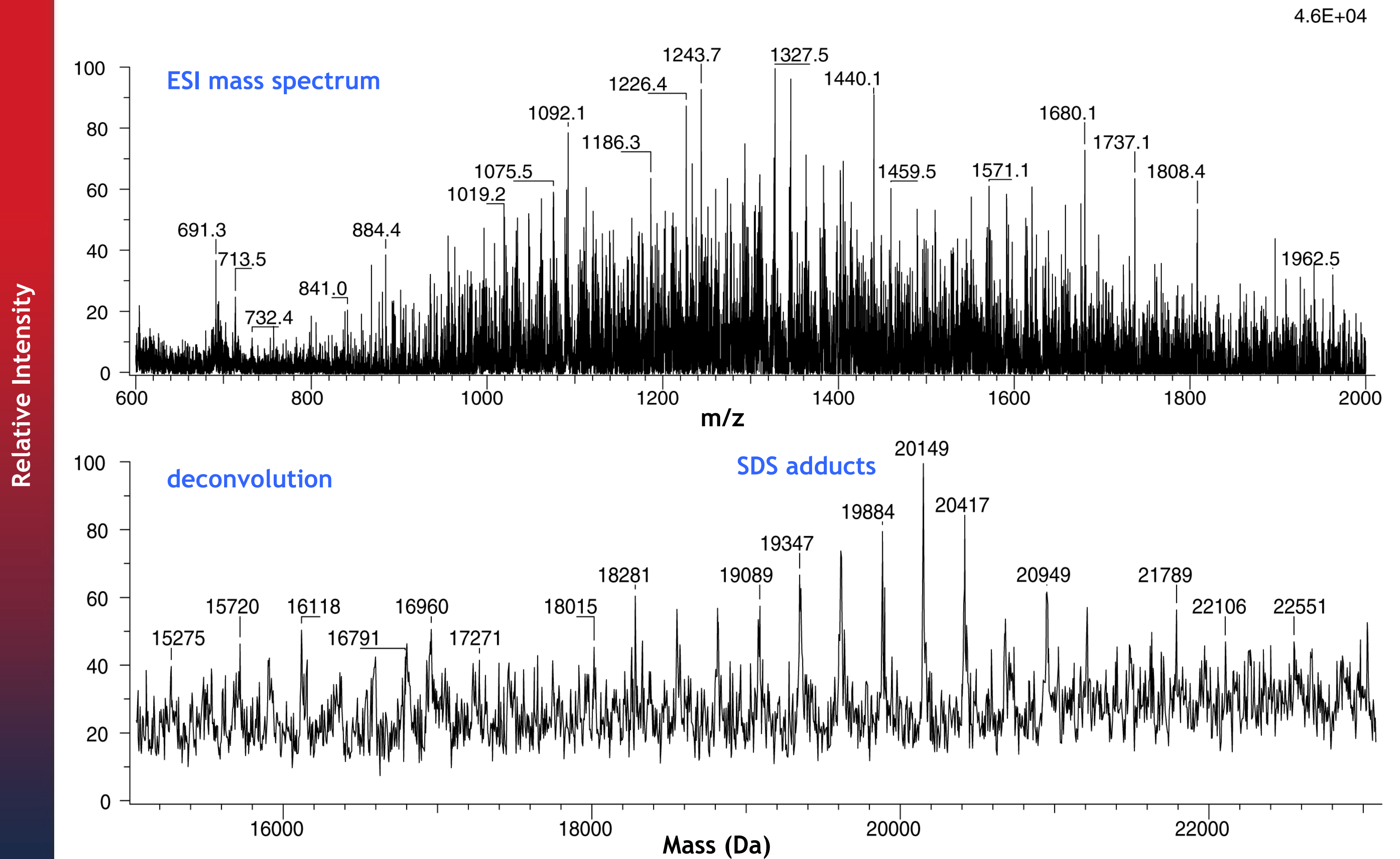
50 pmoles of crude isolate in 300 uL



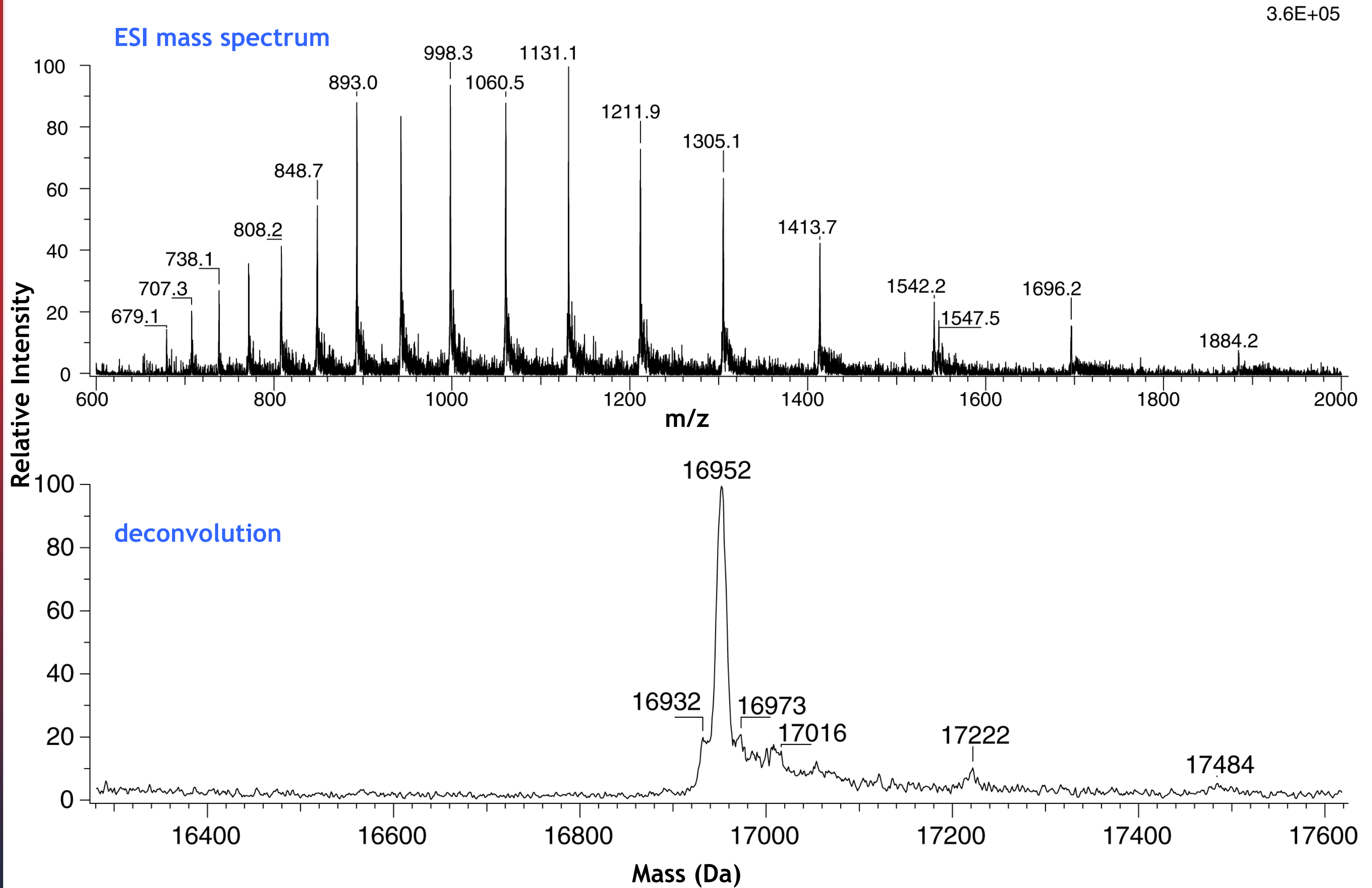
Automated protein sample cleanup system with on-line SDS detergent removal



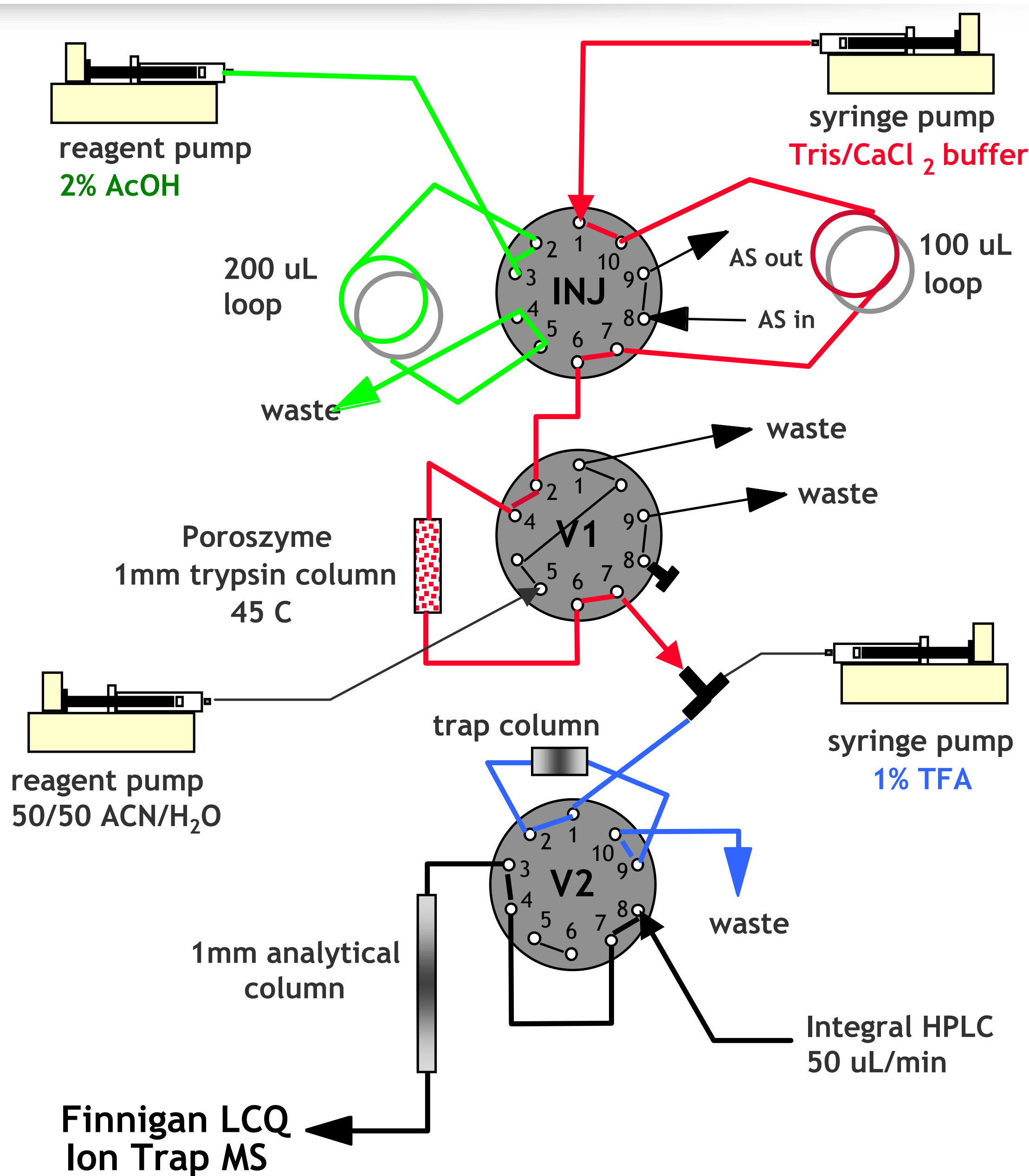
Mass spectra of 25 pmol myoglobin in 0.1% SDS no SDS removal



Mass spectra of 25 pmol myoglobin in 0.1% SDS using **automated SDS removal**



Automated Micro-scale On-line Digest/Ion Trap System



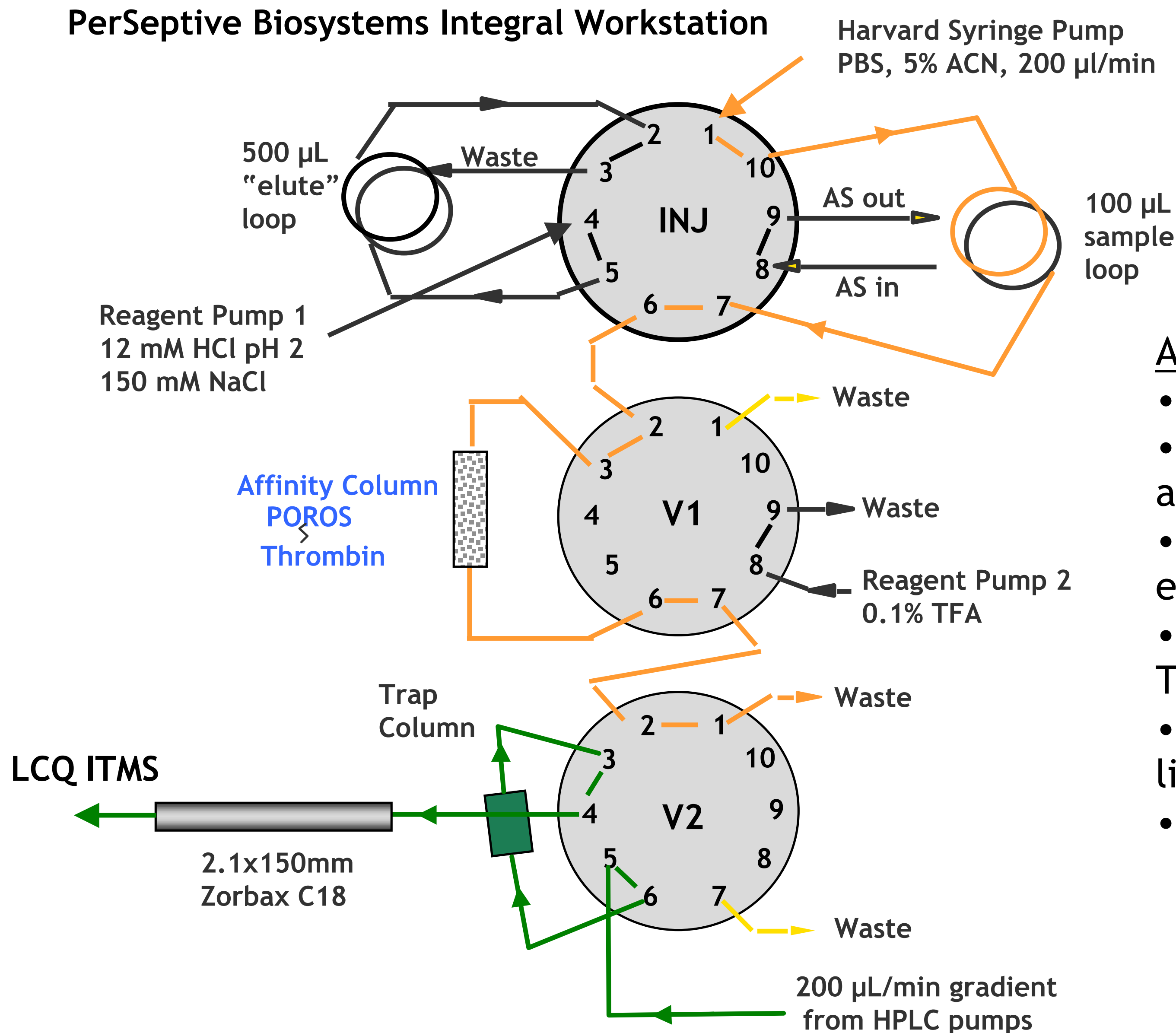
PerSeptive Biosystems Integral Workstation

Digest Sequence

- draw sample into loop
- backflush thru PZT w/digest buffer
- flip INJ, wash PZT, desalt trap
- flip V2, gradient elute peptides
- flip V1, clean PZT

Schematic of On-line Affinity LC/MS System

"automated affinity applause meter"

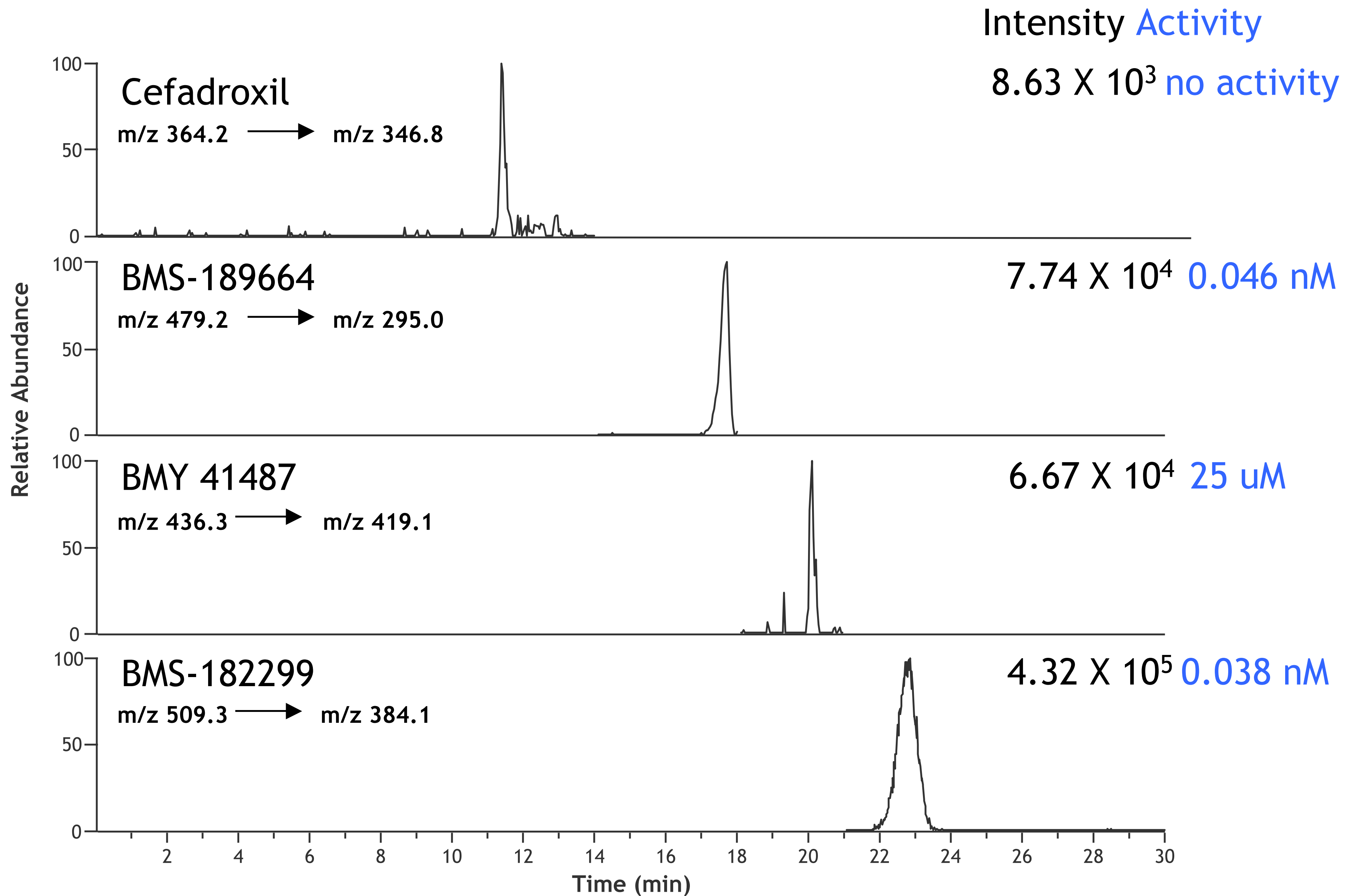


Affinity LC/MS sequence:

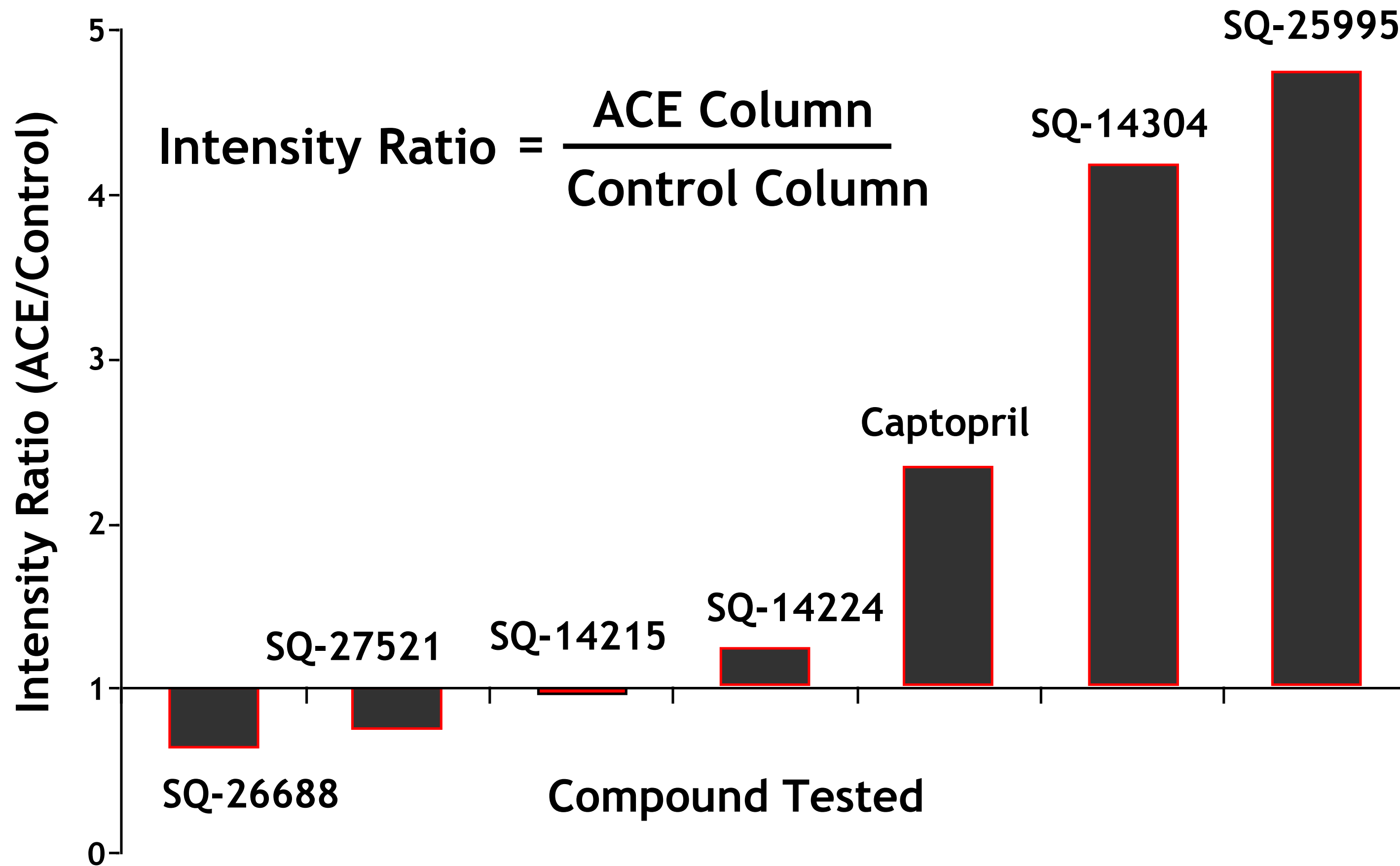
- draw sample into loop
- backflush, capture ligands on affinity column
- flip V2 (trap in line), flip INJ, elute ligands onto trap
- flip V1, wash trap with 0.1% TFA
- flip V2, elute/separate ligands
- detect by LC/MS^n

Affinity LC/Ion Trap/MS/MS of 4-Component Mixture

using immobilized Thrombin



Affinity LC/MS Binding Profile of ACE Inhibitors



In Vitro Activity

SQ-25995	> Captopril	> SQ-14304	> SQ-14224	> SQ14215	> SQ27521	> SQ26688
2.9 nM	23 nM	435 nM	965 nM	5.7 M	565 M	2.4 mM

Conclusions: Analytical Figures of Merit of LC/MS for Biomolecule Analysis

- Excellent Mass Accuracy
  - +/- 0.01% routine (1 Da in 10 kDa)
- Resolving Power
  - e.g., a 100 Da change can be detected in a 150 kDa antibody
- Rapid
  - minutes
- High Sensitivity
  - picomole – attomole depending on sample handling
- Automation capability
  - increased throughput
  - increased sensitivity via reduction of manual sample handling steps
- LC/MS^N is a powerful tool for biomolecule structure analysis, -profiling, sequence, interaction

Acknowledgments

- Bristol-Myers Squibb
  - Mark Sanders, Mike Nedved, Haiying Zhang, Subinay Ganguli
- Tufts University
  - Dr. Dan Jay

Internet Resources for LC/MS and Bio-MS

- ThermoFinnigan
  - http://www.thermofinnigan.com
- Novatia (brand new but adding new content – tips on SEQUEST usage)
  - http://www.enovatia.com
- Base peak from Wiley (a good MS link site, commercial and academic)
  - http://base-peak.wiley.com
- ABRF (good source for biological apps, join the ABRF list server)
  - http://www.abrf.org
- Ion Source (good information and tutorials, links)
  - http://ionsource.com
- Prowl at Rockefeller University (good bioMS site, useful software - PAWS)
  - http://prowl.rockefeller.edu