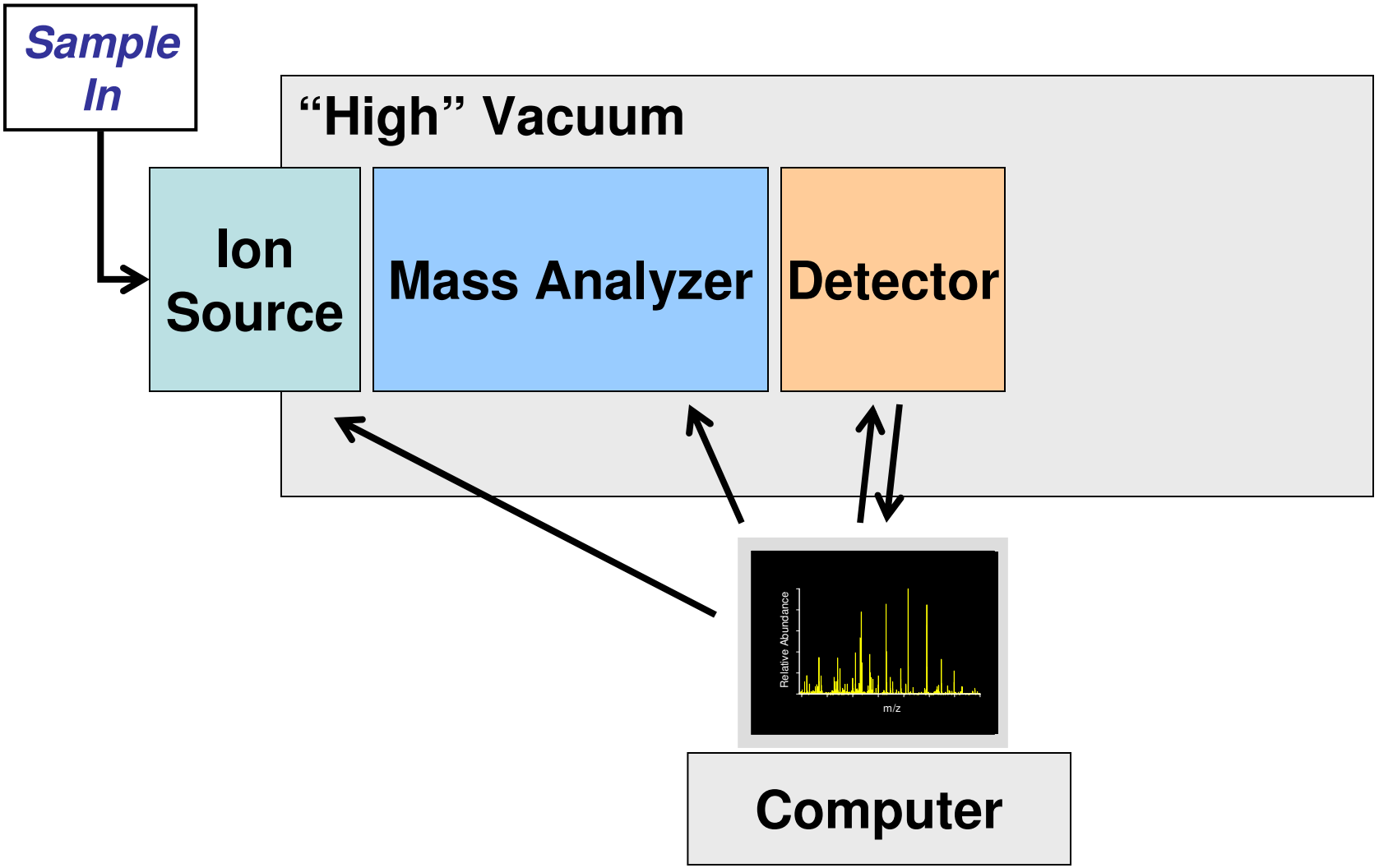
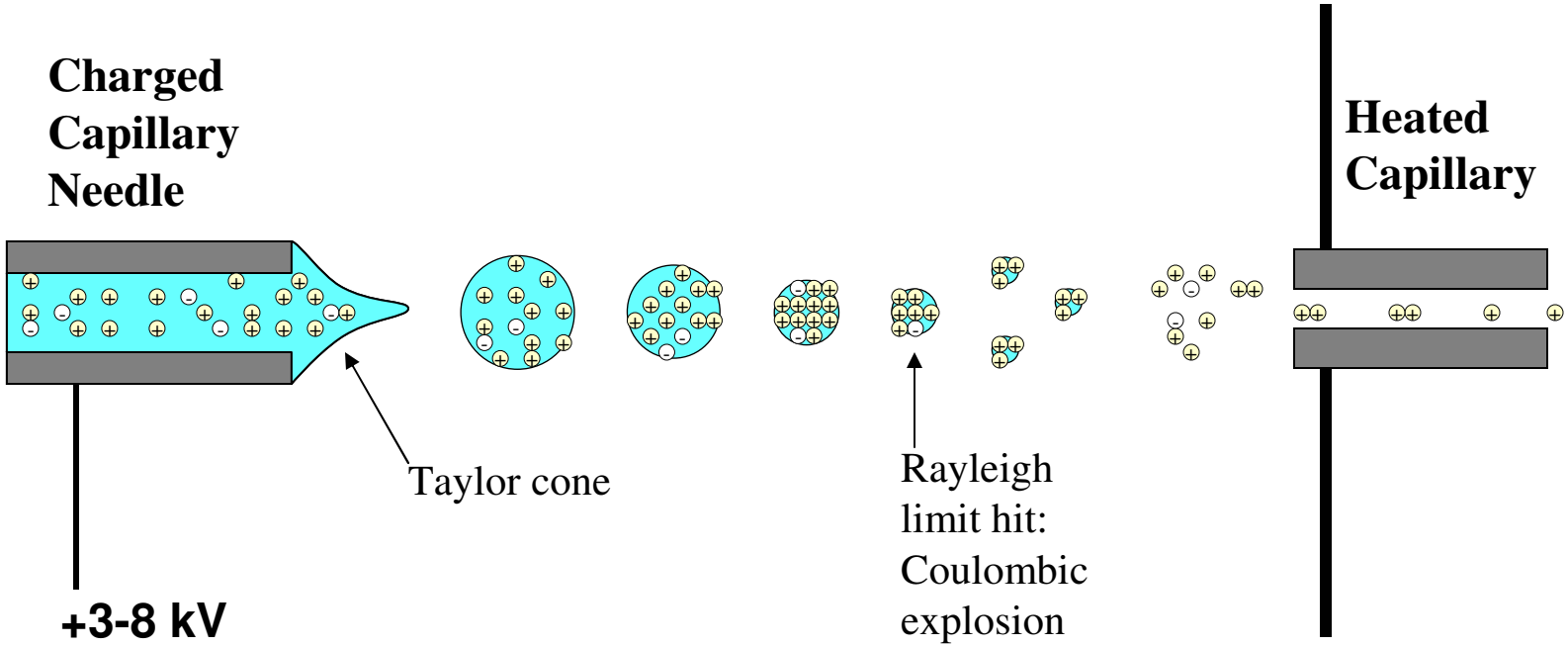
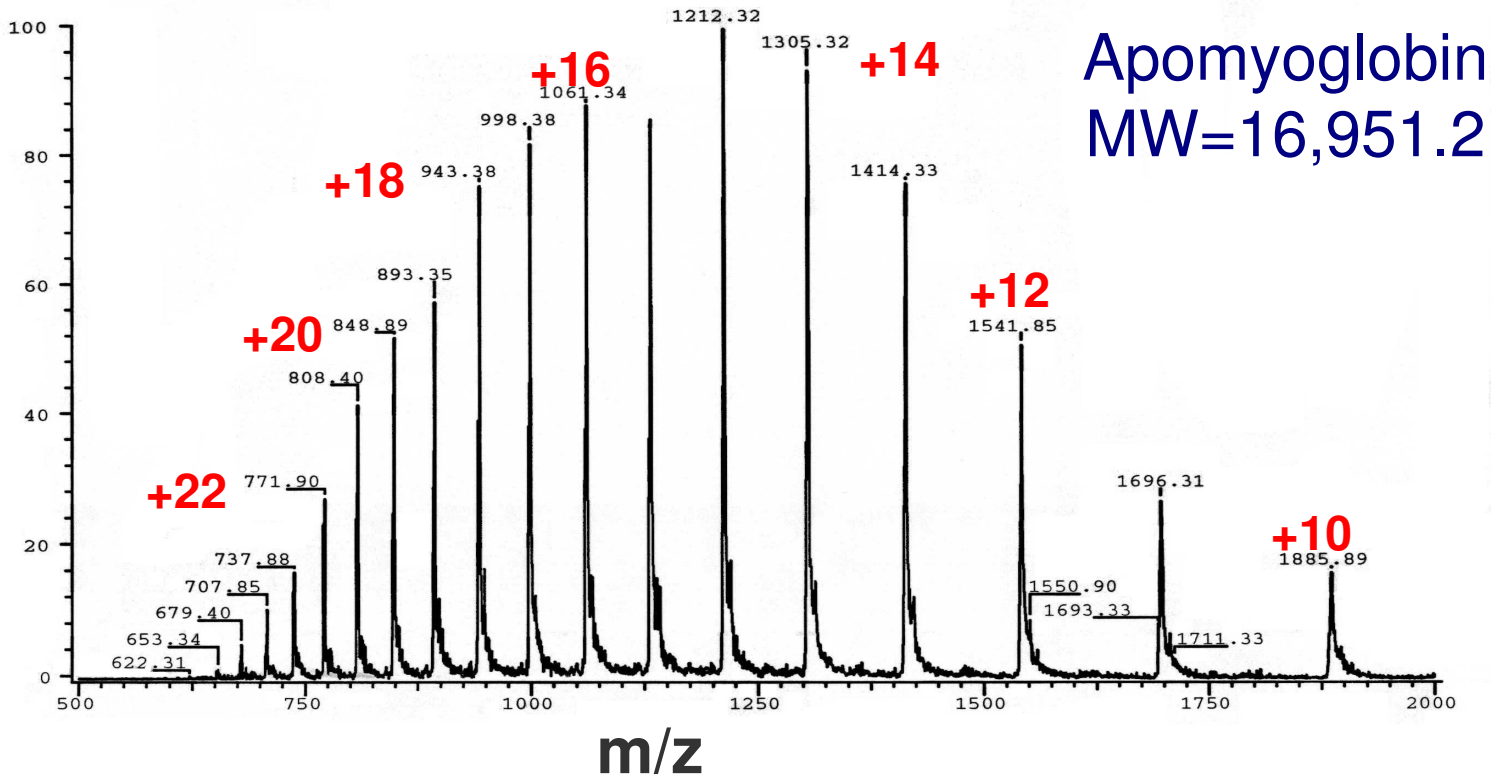
Block Diagram of a Mass Spectrometer



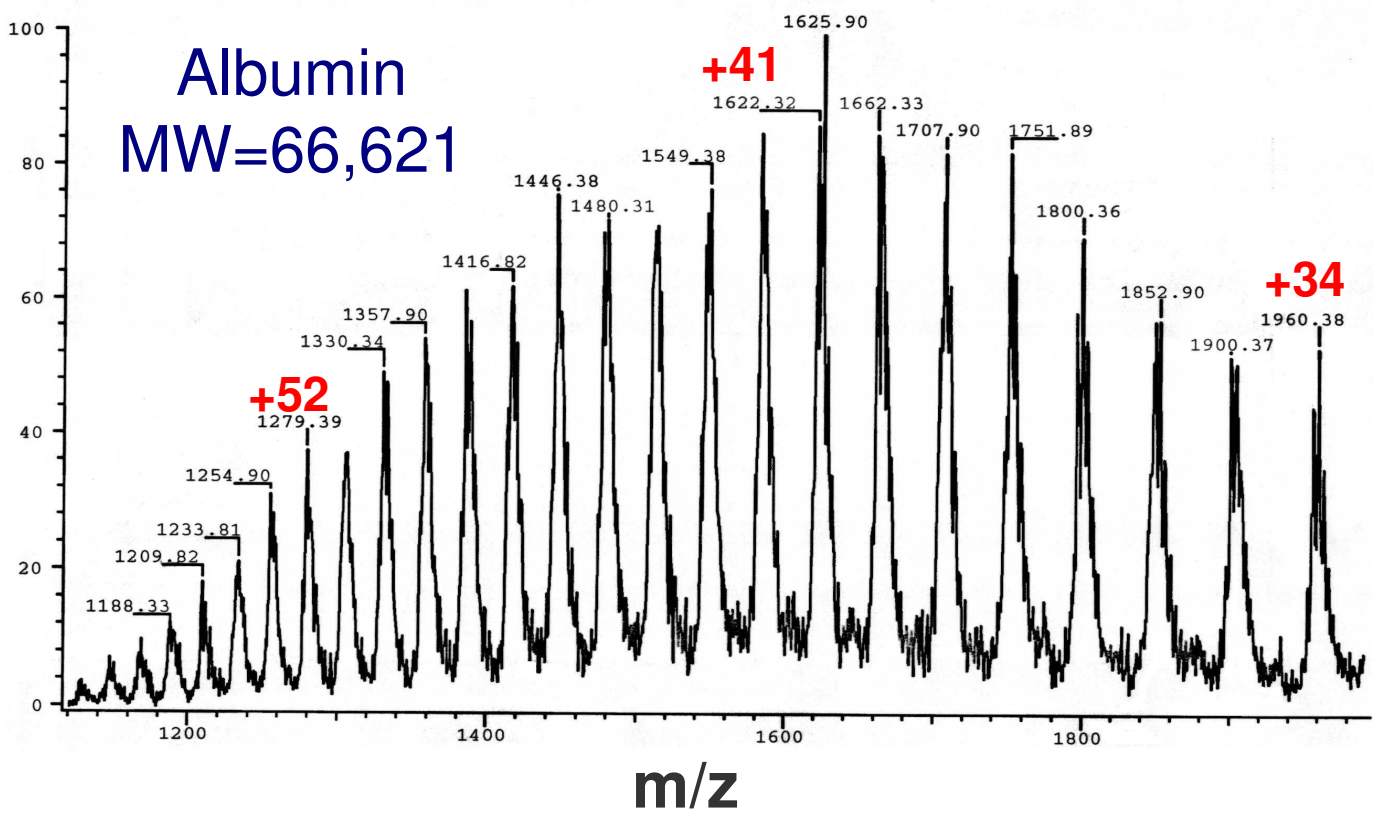
Mechanism of Electrospray Ionization (ESI)



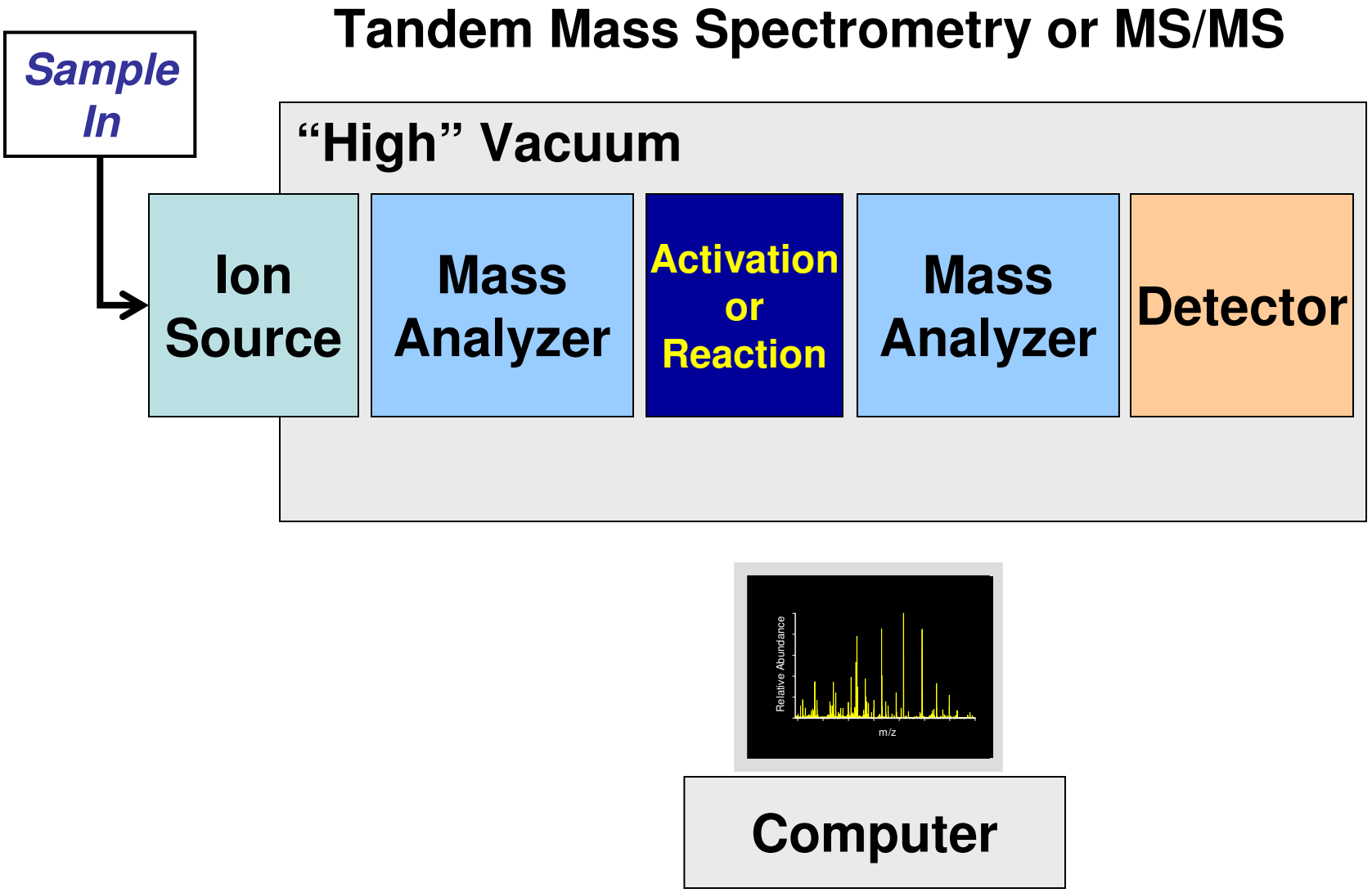
ESI-MS Spectrum of Horse Apomyoglobin



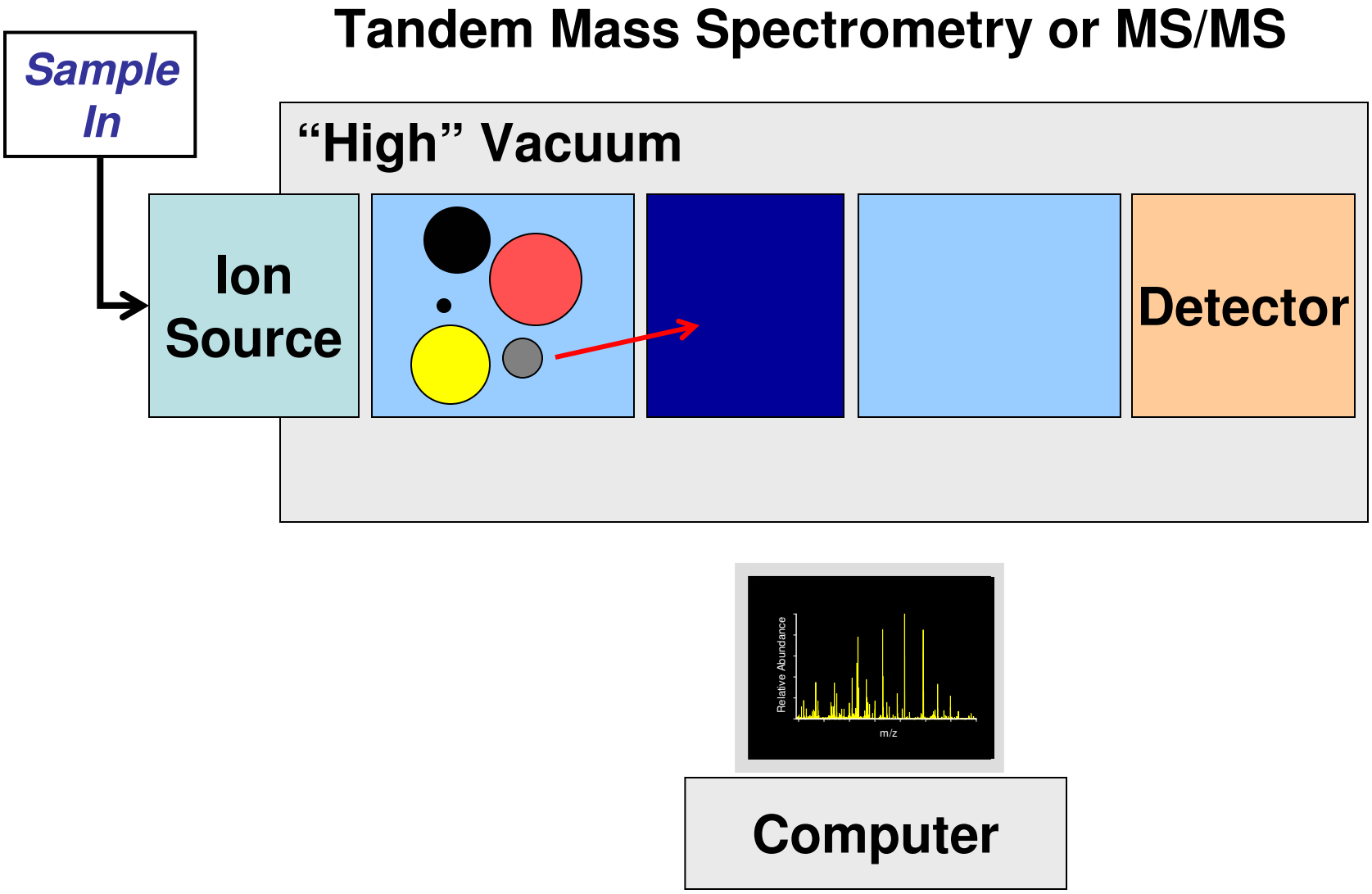
ESI-MS Spectrum of Bovine Serum Albumin



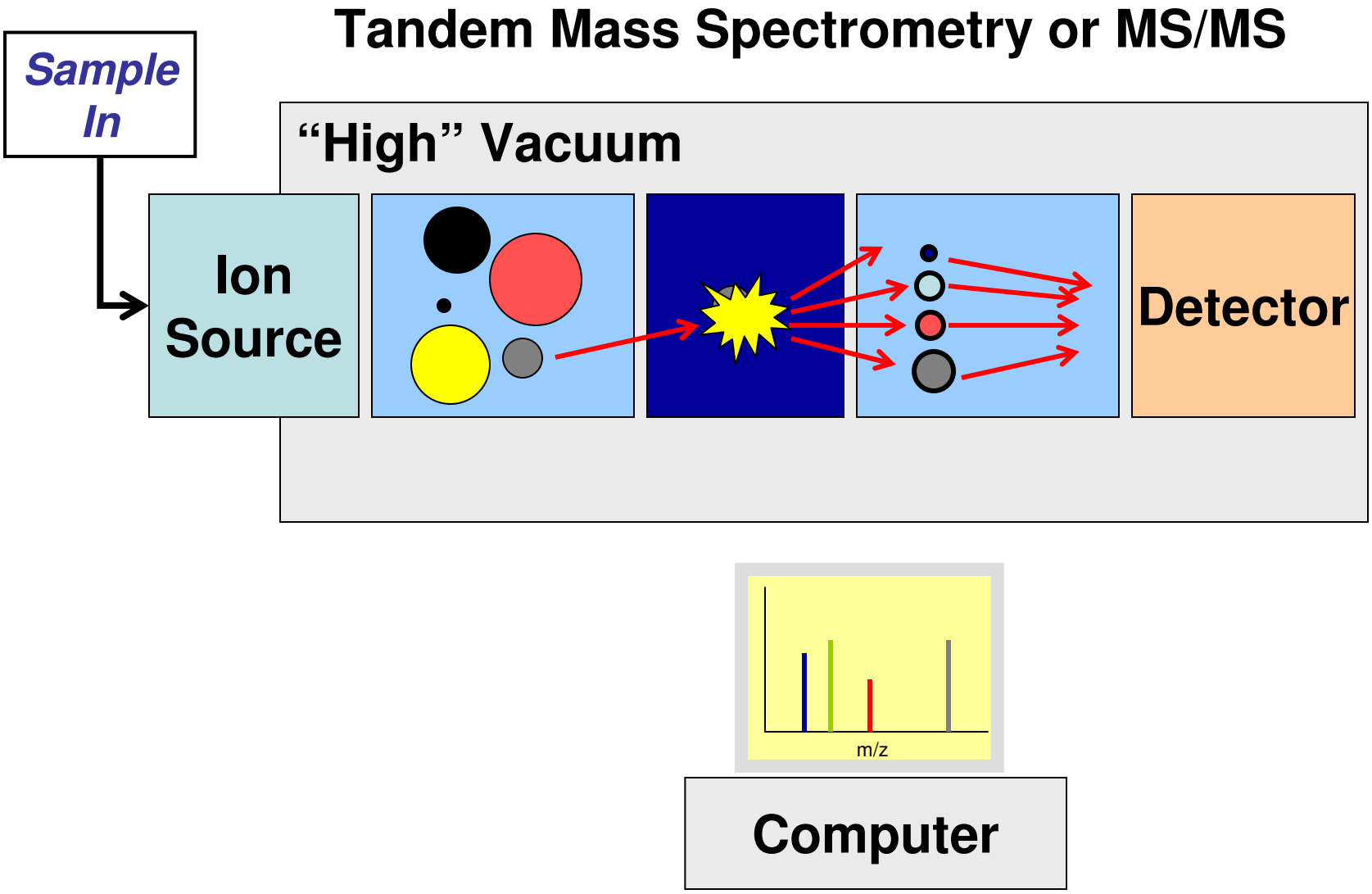
Block Diagram of a Mass Spectrometer



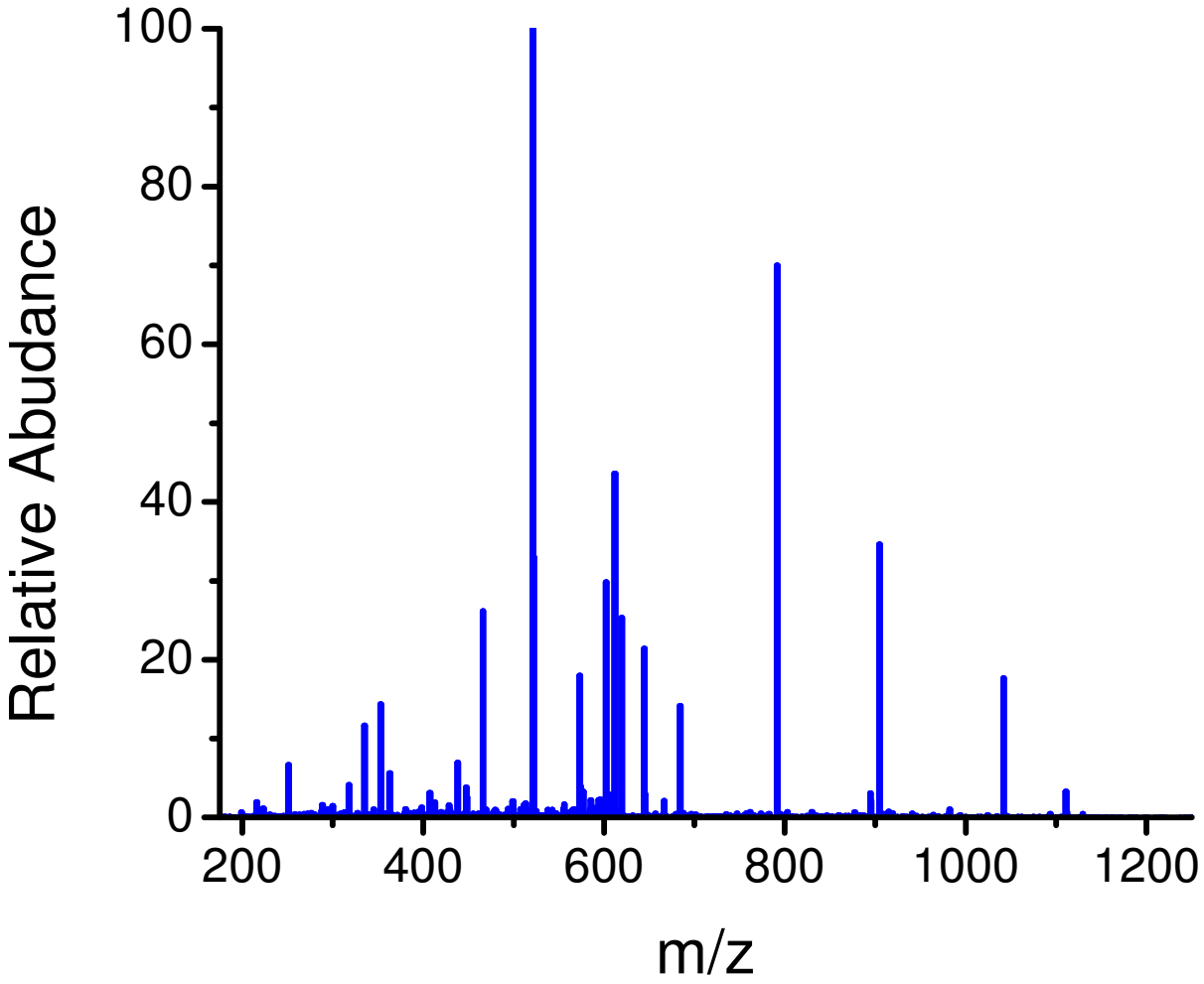
Block Diagram of a Mass Spectrometer

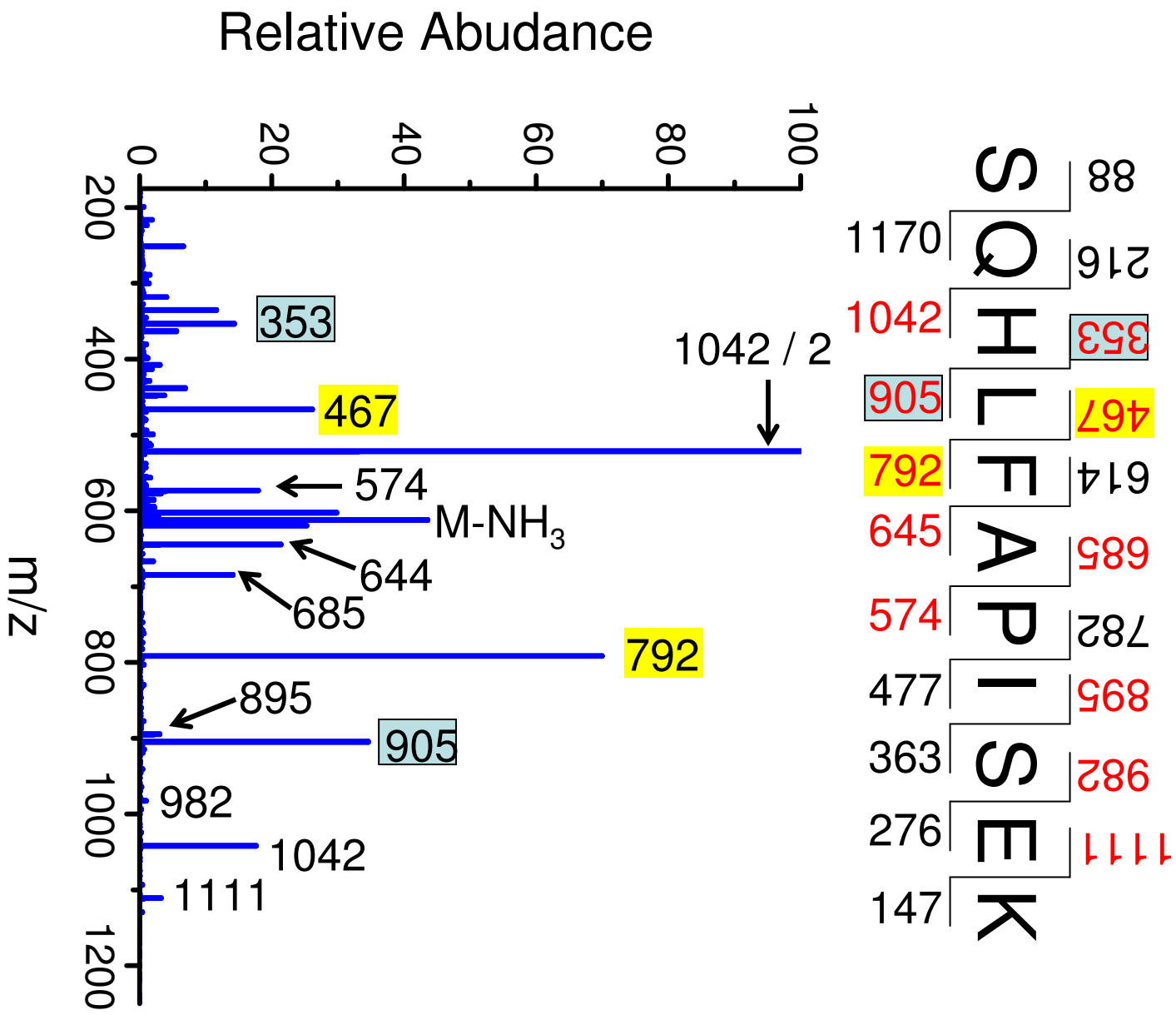


Block Diagram of a Mass Spectrometer

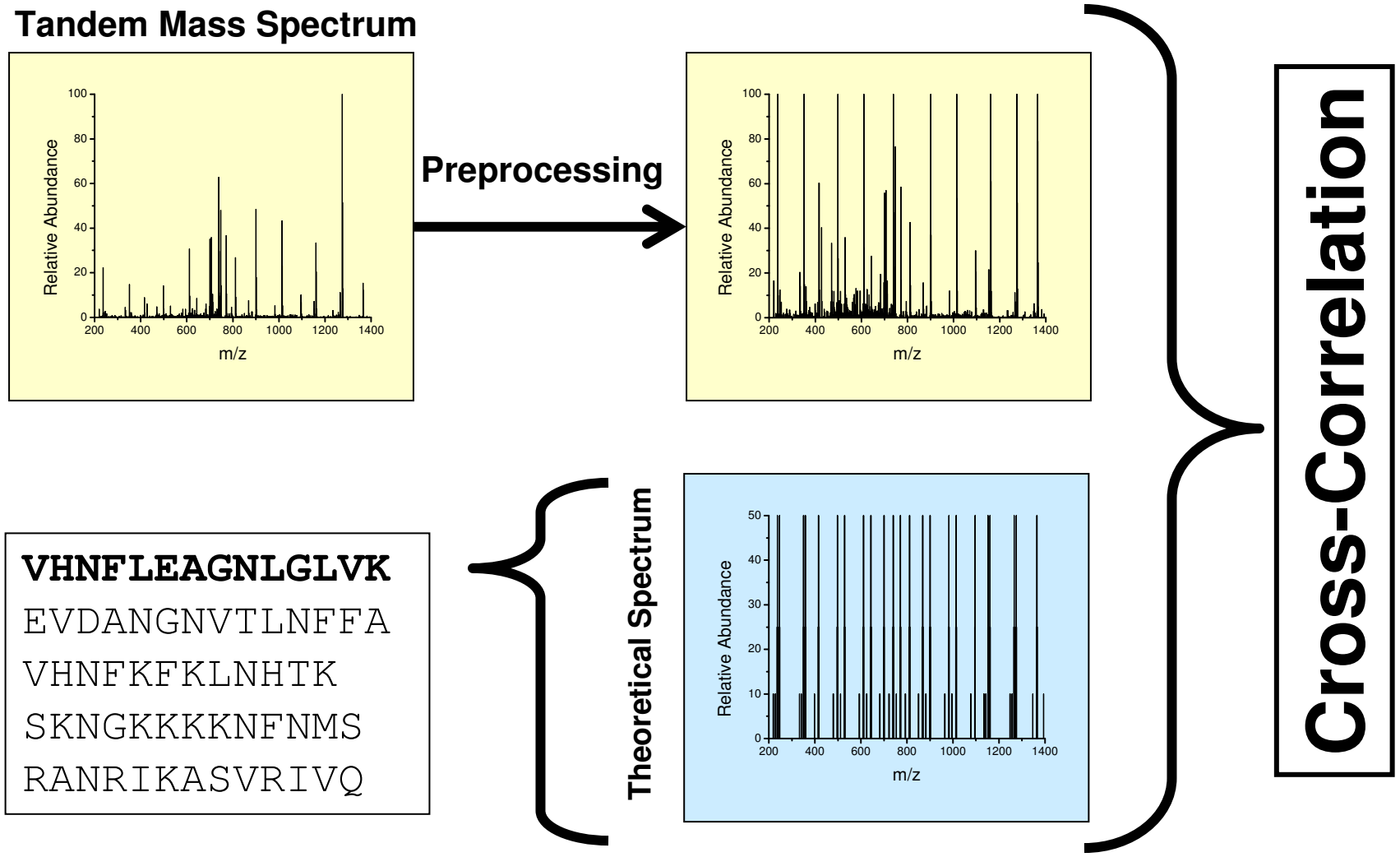


S Q H L F A P I S E K

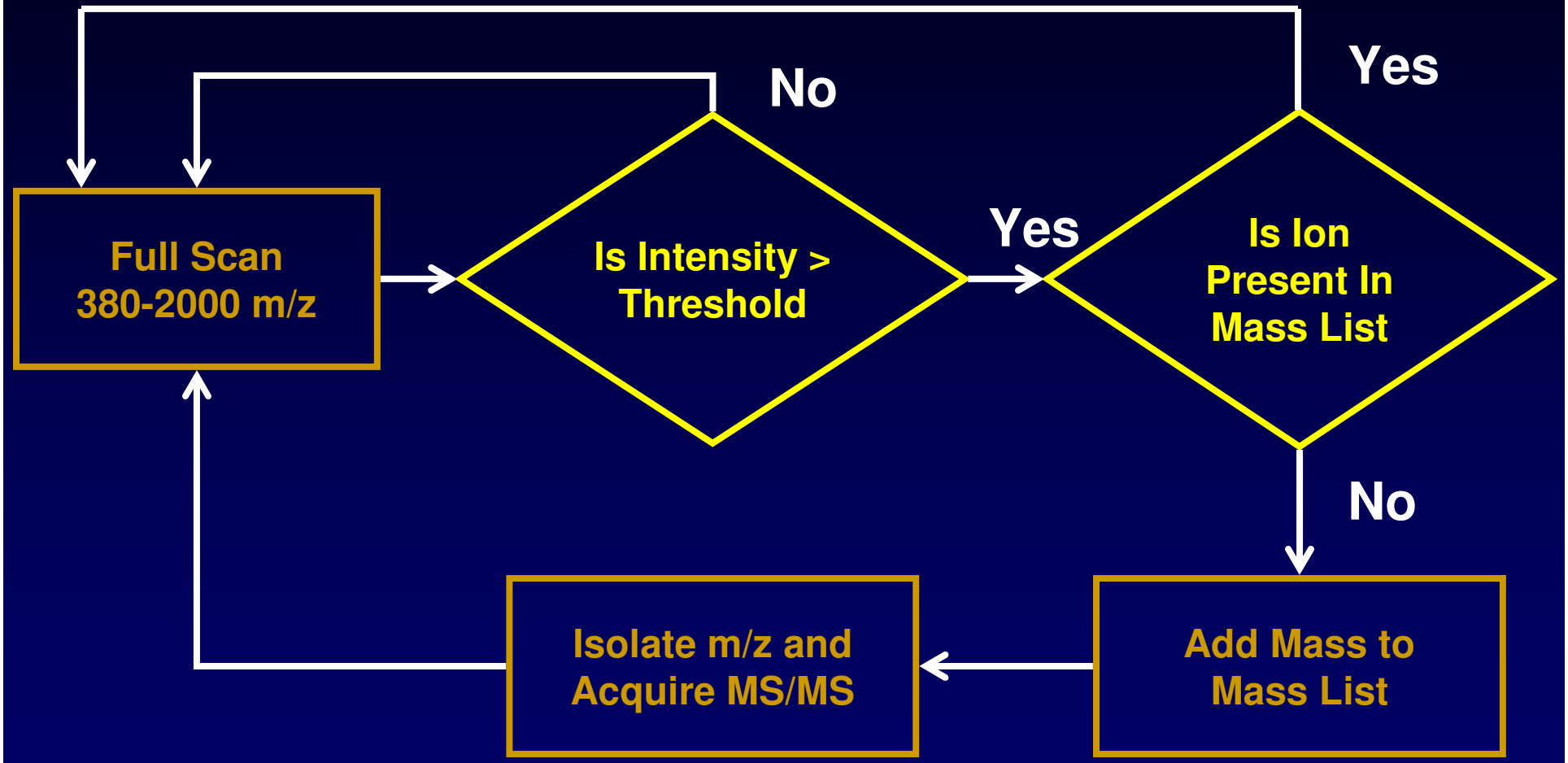



Relative Abudance
0
20
40
60
80
100
m/z
200
400
600
800
1000
1200
353
467
1042 / 2
574
M-NH3
644
685
792
895
905
982
1042
1111
S
Q
H
L
E
A
P
I
S
E
K
88
1170
216
1042
353
905
467
792
614
645
685
574
782
477
895
363
982
276
1111
147

Identification of Peptide Sequences Using Tandem Mass Spectra and SEQUEST



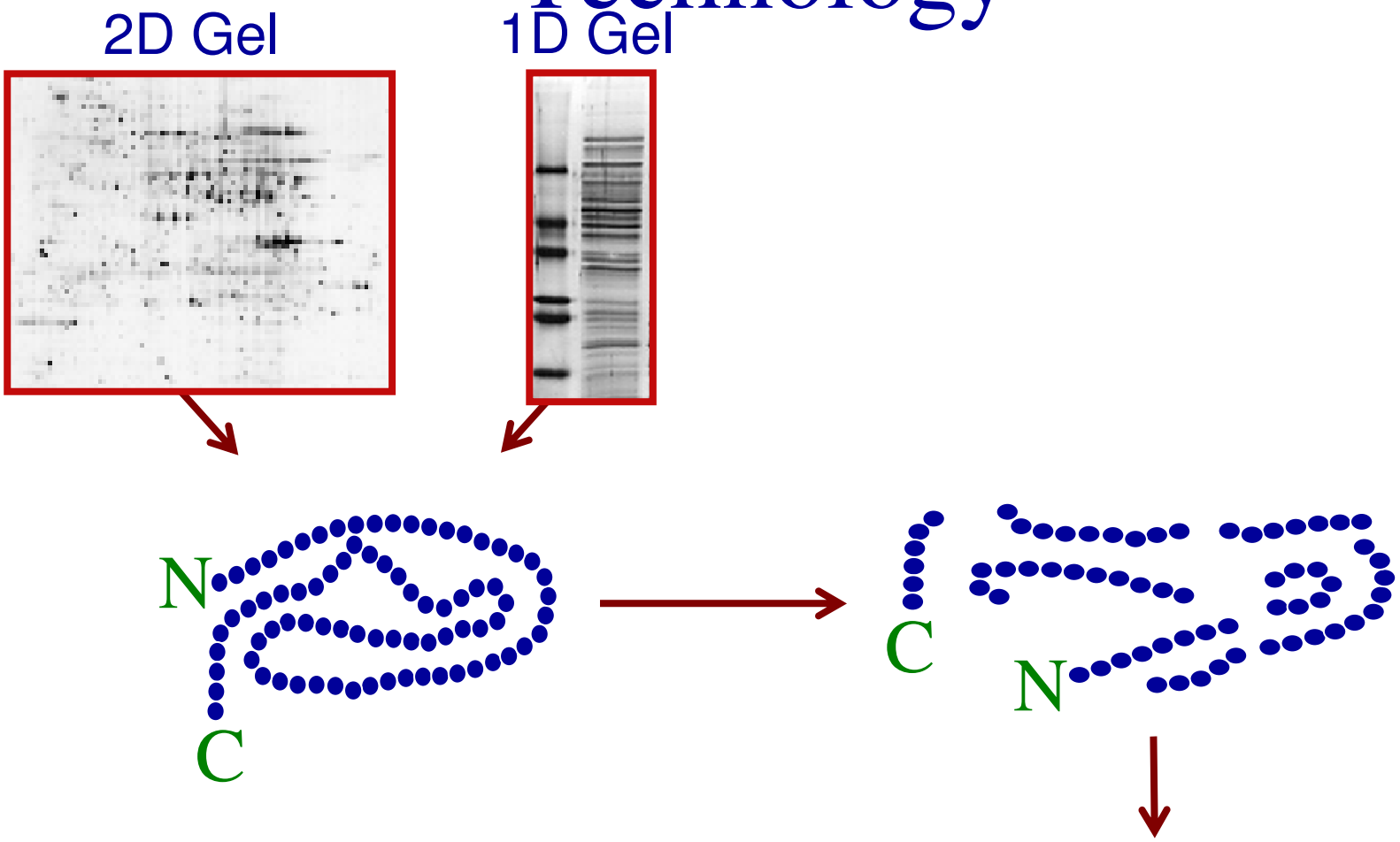
Data Dependant Acquisition



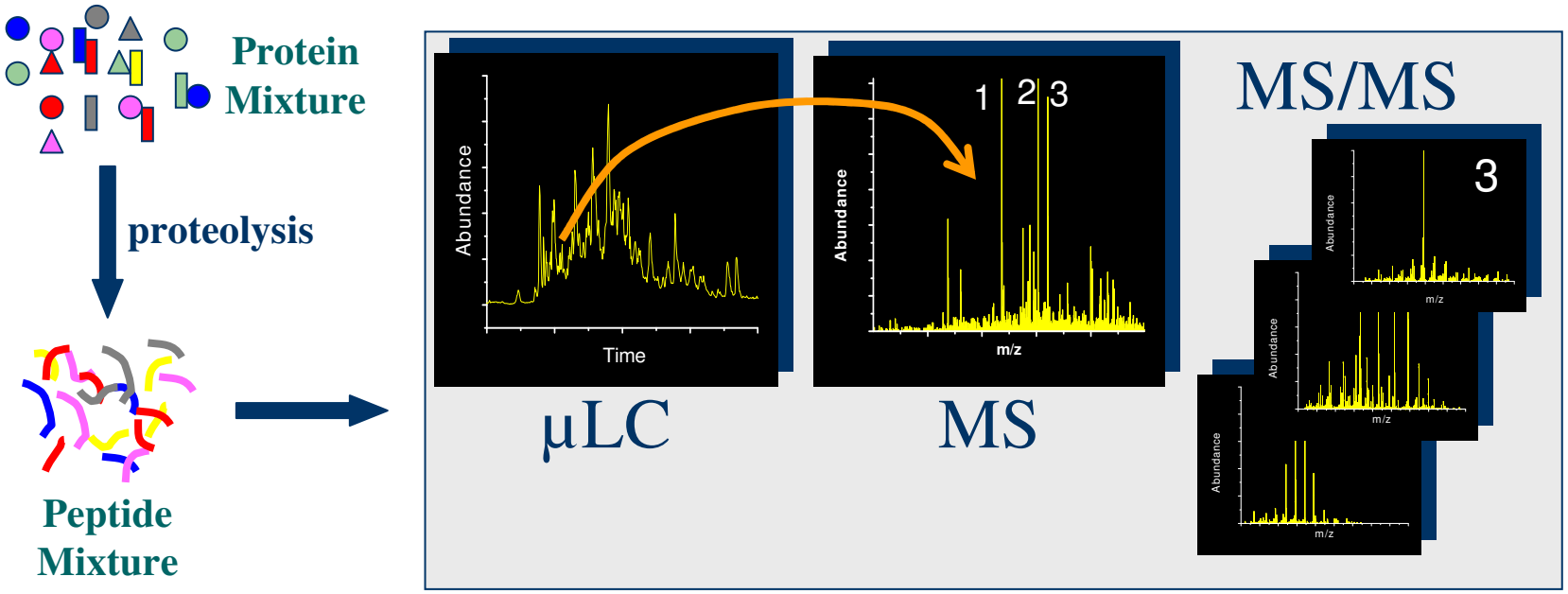
Yates *et al.* Anal Chem, 67, 1426 (1995)

Stahl *et al.* JASMS, 7, 532 (1996)

Proteomics: Traditional Technology

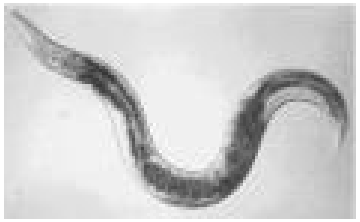


Shotgun Proteomics

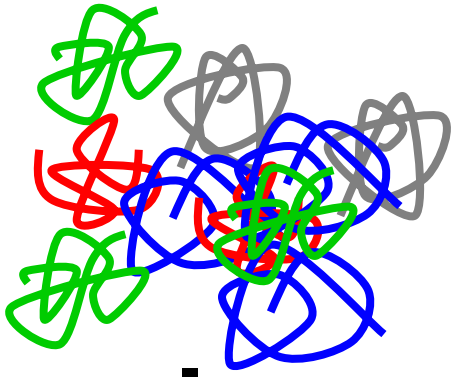


Shotgun Characterization of Protein Mixtures

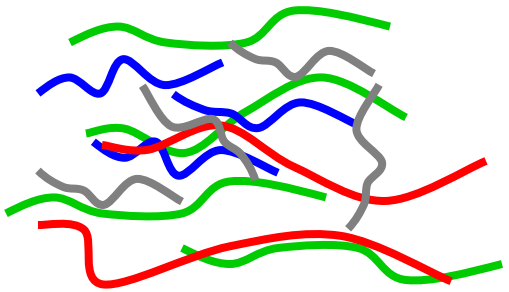
C. elegans



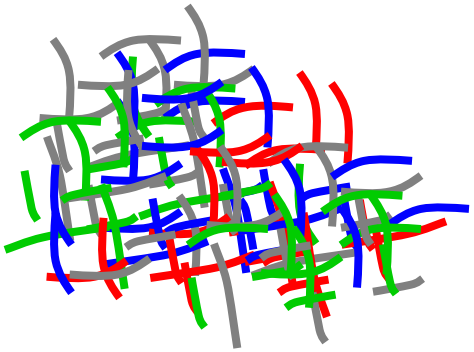
Extract Protein Mixture



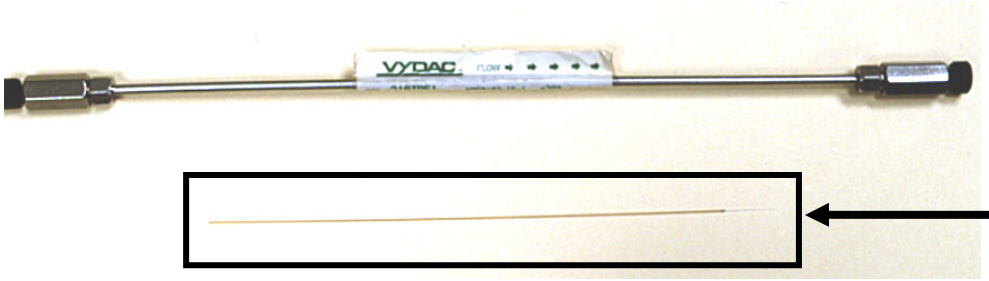
Denature



Digest into Small Peptides

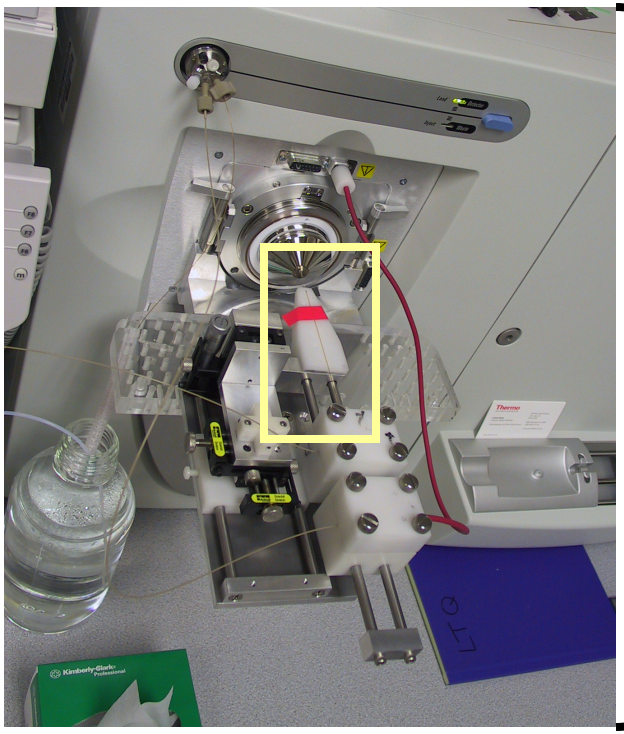


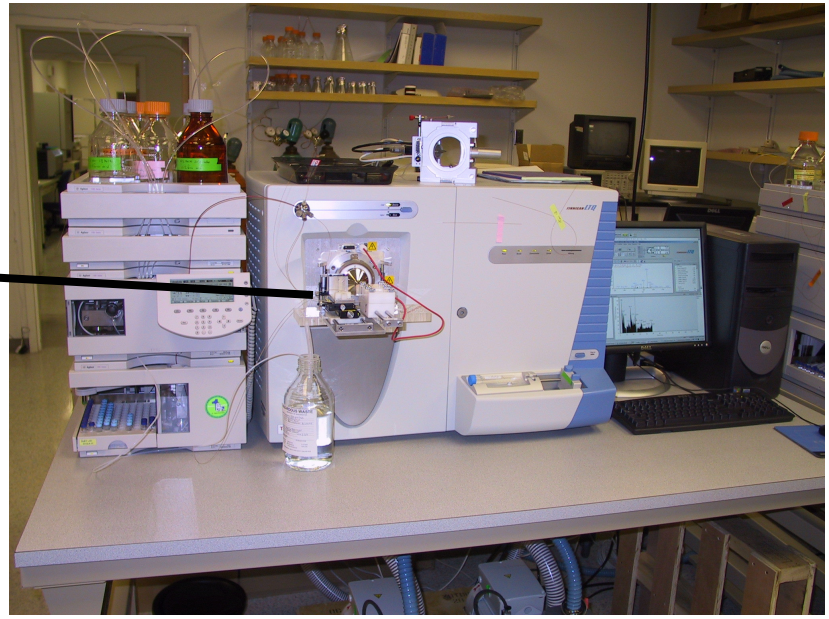
Measurement of Peptides by Microcapillary Liquid Chromatography Tandem Mass Spectrometry



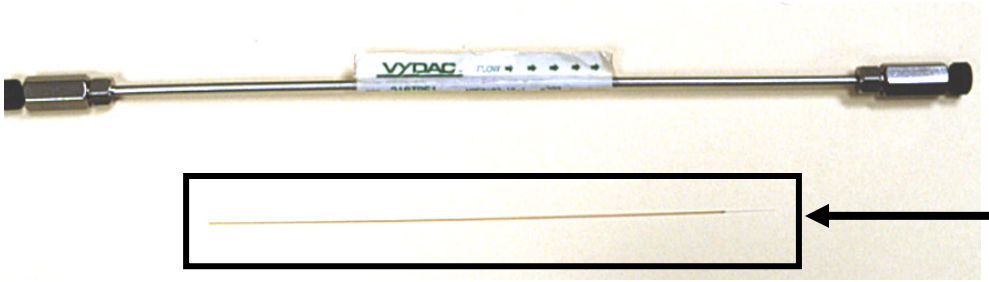
Traditional Chromatography Column

Microcapillary Chromatography Column



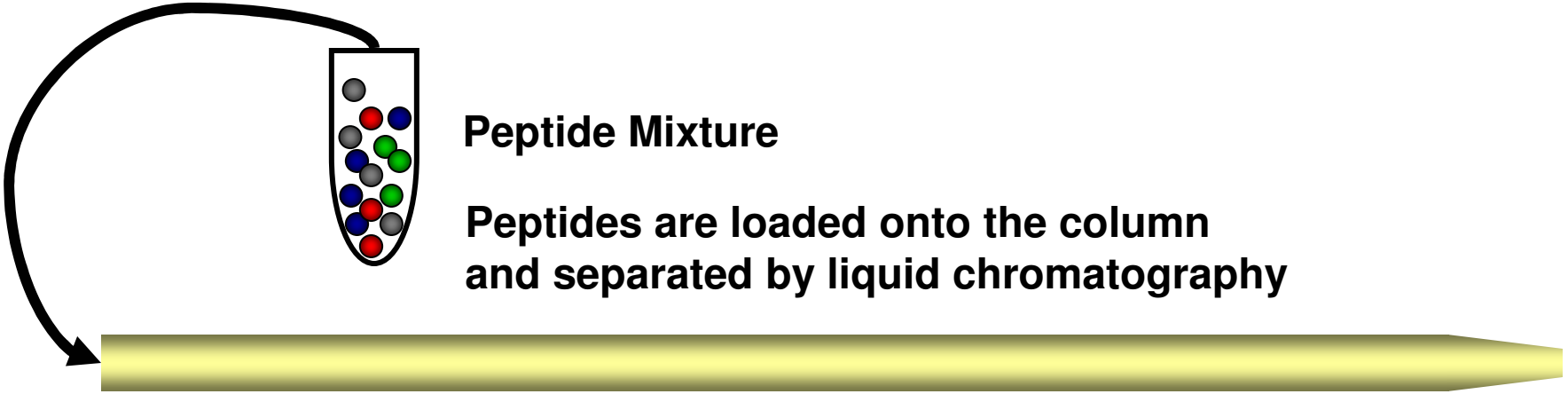


Measurement of Peptides by Microcapillary Liquid Chromatography Tandem Mass Spectrometry



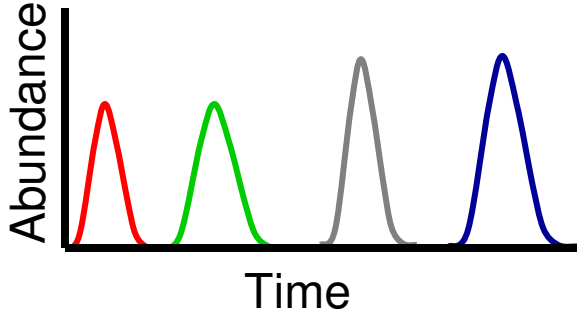
Traditional Chromatography Column

Microcapillary Chromatography Column

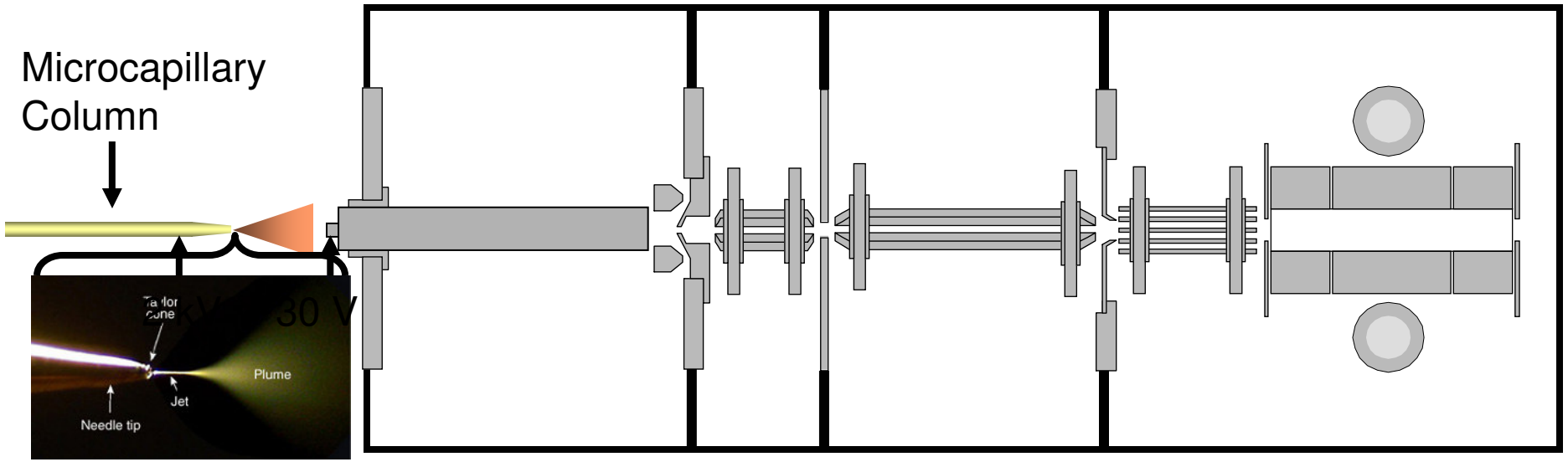


Peptide Mixture

Peptides are loaded onto the column and separated by liquid chromatography

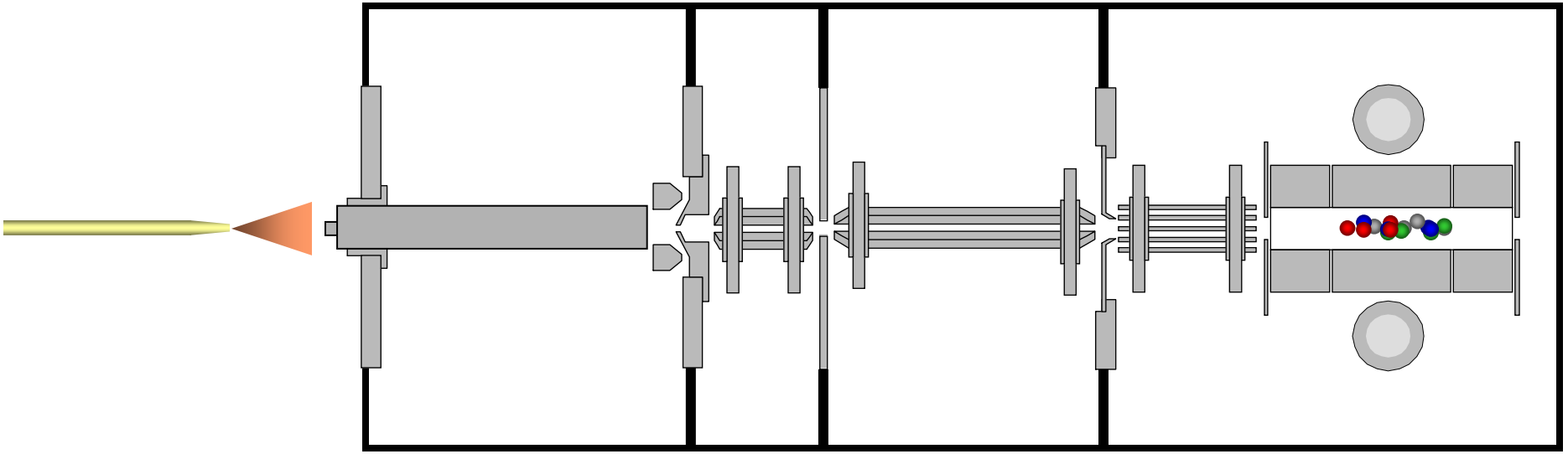


Acquiring Mass Spectrometry Data of Peptides

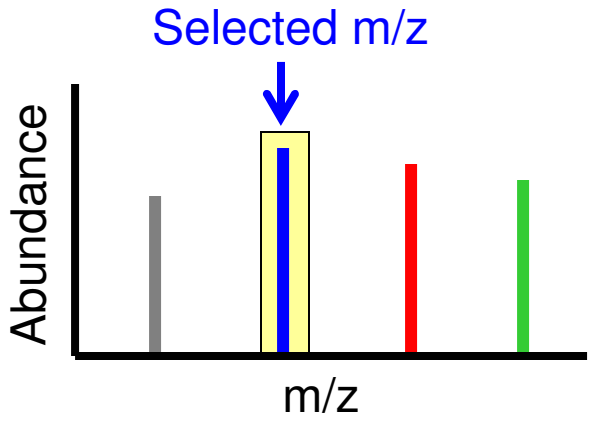


- Peptides from the column are ionized and transmitted into the vacuum system using electrospray and trapped in the 2-dimensional ion trap

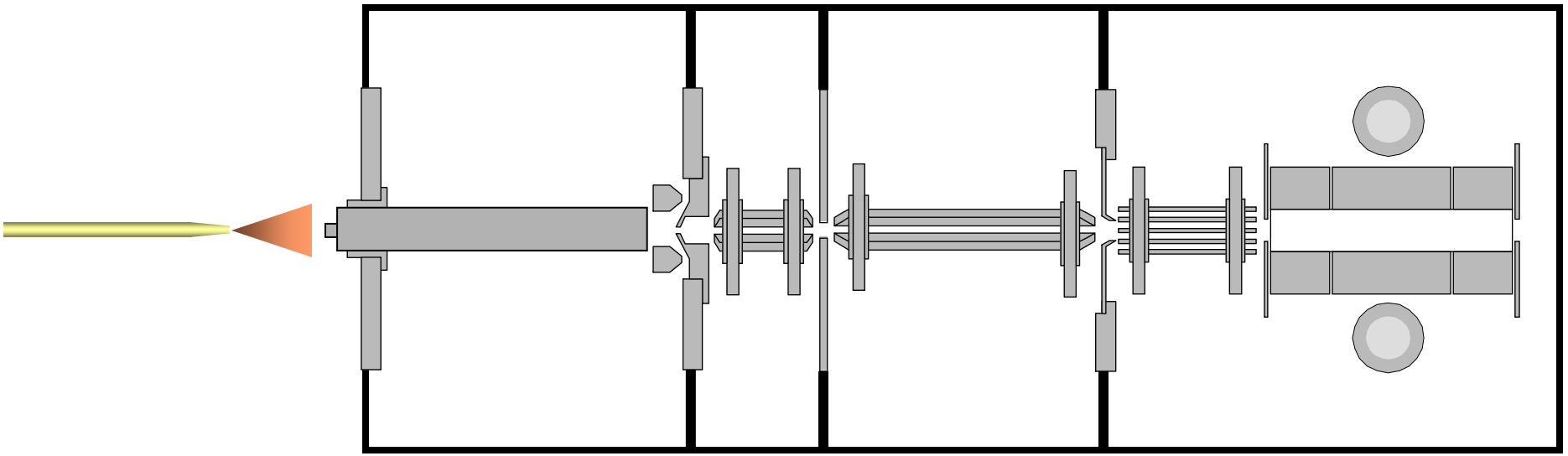
Acquiring Mass Spectrometry Data of Peptides



- Peptides from the column are ionized and transmitted into the vacuum system using electrospray and trapped in the 2-dimensional ion trap
- Ions are selectively ejected radially by their m/z to produce a mass spectrum
- The instrument data-system automatically selects a m/z for further analysis

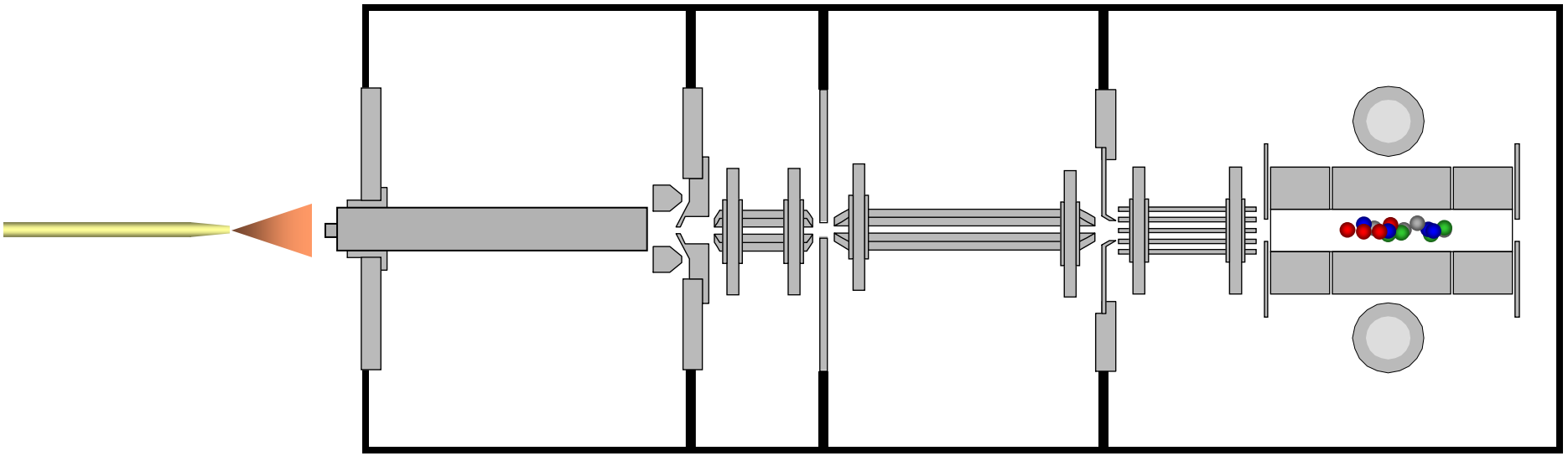


Acquiring a Fragmentation Spectrum



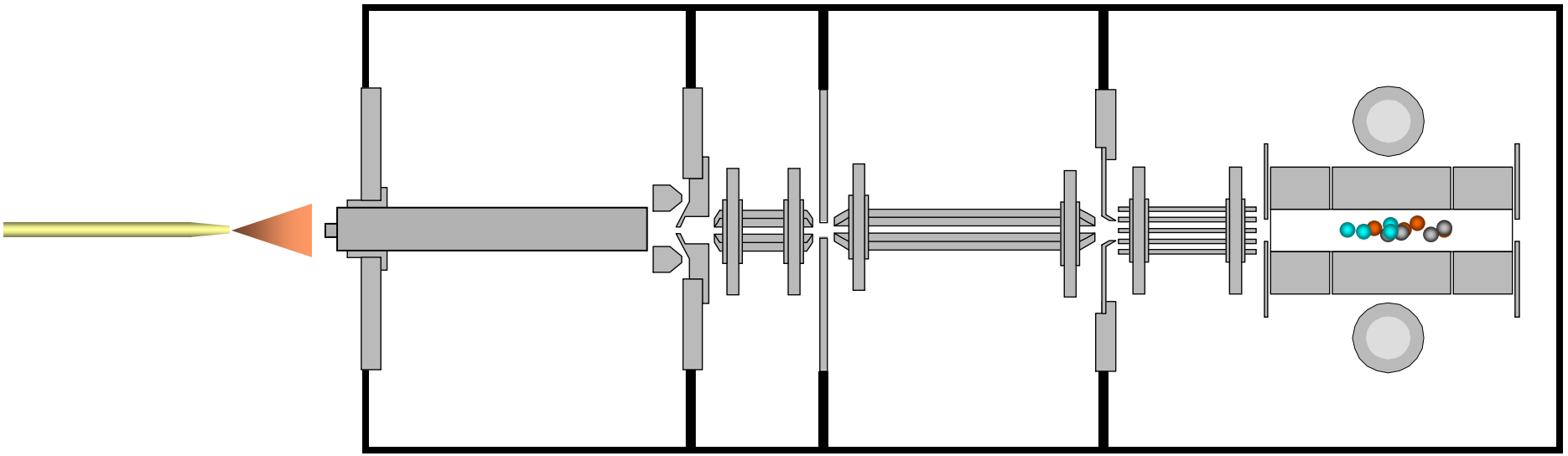
- The trap is then refilled

Acquiring a Fragmentation Spectrum

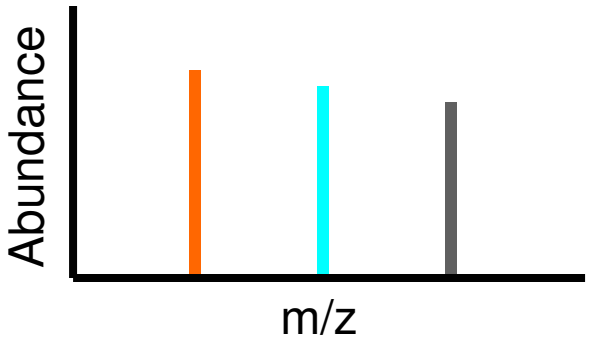


- The trap is then refilled
- All ions are ejected from the trap except the predetermined m/z
- The energy of the isolated ion is increased resulting in collisions with the helium gas in the trap

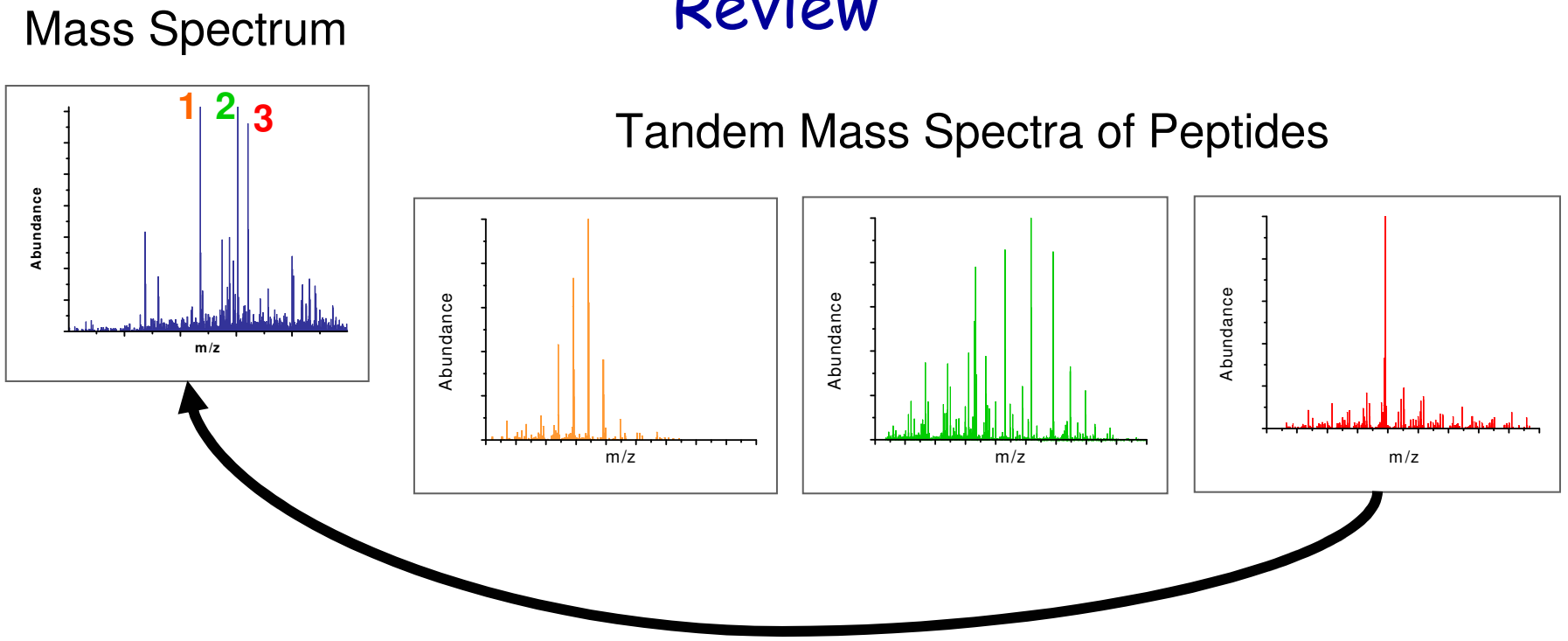
Acquiring a Fragmentation Spectrum



- The collisions with the helium molecules result in structure specific fragment ions
- The fragment ions are then selectively ejected to produce a tandem mass spectrum



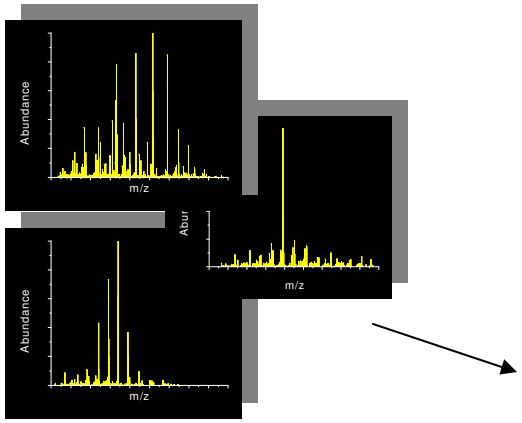
Review



1 to 2 Seconds per cycle

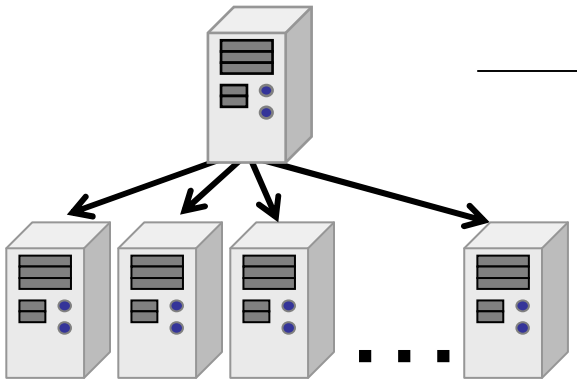
1. Acquire mass spectrum
2. Select precursors for tandem mass spectrometry
3. Selectively isolate and fragment precursor ions.
4. Repeat for 3 different precursor ions

Spectrum Identification



MS/MS spectra

SEQUEST



Beowulf Computer Cluster

Peptide ID list

```
Scan1 0.7 EGSSDEEVP…
Scan1 0.3 TFAEILNPI…
Scan1 0.2 ARFDLNNHD…
 ------------------
Scan2 0.5 EDEESIRAV…
Scan2 0.2 WLGDDCFMV…
Scan2 0.1 IDRAAWKAV…
 ------------------
Scan3 0.2 EITTRDMGN…
Scan3 0.1 GRNMCTAKL…
```

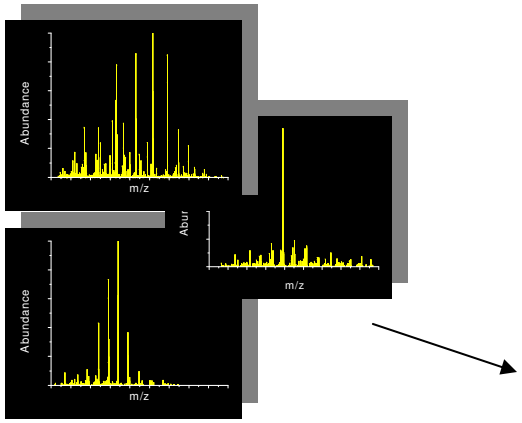
>MEKK1 (kinase)
MDRILARMKKSTRRGGDKNI
TPVRRLERR…

>ATMKK5 (kinase kinase)
MKPIQSPSGVASPMKNRLRK
RPDLSPPLPHRDVALAVLP…

Database: fasta file

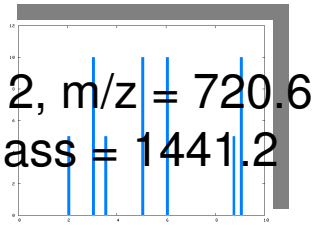
ID proteins from peptides…

Spectrum Identification



MS/MS spectra

SEQUEST



z = 2, m/z = 720.6
mass = 1441.2

score = 0.3

Peptide ID list

```
Scan1  0.7  EGSSDEEVP…
Scan1  0.3  TFAEILNPI…
Scan1  0.2  ARFDLNNHD…
 ------------------
Scan2  0.5  EDEESIRAV…
Scan2  0.2  WLGDDCFMV…
Scan2  0.1  IDRAAWKAV…
 ------------------
Scan3  0.2  EITTRDMGN…
Scan3  0.1  GRNMCTAKL…
```

>MEKK1 (kinase)
MDRILARMKKSTRRGGDKNI
TPVRRLERR…

>ATMKK5 (kinase kinase)
MK TFAEILNPITWR LR
KRPDLSPPLPHRDVALAVP…

Database: fasta file

Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry

Matthias Wilm*, Andrej Shevchenko*, Tony Houthaeve*, Stephen Breit†, Lothar Schweigerer†‡, Theodore Fotsis†‡ & Matthias Mann*§

*Protein & Peptide Group, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
†Division of Oncology and Haematology, Children's Hospital, University of Heidelberg, INF 150, 69120 Heidelberg, Germany

Molecular analysis of complex biological structures and processes increasingly requires sensitive methods for protein sequencing. Electrospray mass spectrometry[1] has been applied to the high-sensitivity sequencing of short peptides[2], but technical difficulties have prevented similar success with gel-isolated proteins. Here we report a simple and robust technique for the sequencing of proteins isolated by polyacrylamide gel electrophoresis, using nano-electrospray[3,4] tandem mass spectrometry[5,6]. As little as 5 ng protein starting material on Coomassie- or silver-stained gels can be sequenced. Multiple-sequence stretches of up to 16 amino acids are obtained, which identify the protein unambiguously if already present in databases or provide information to clone the corresponding gene. We have applied this method to the sequencing and cloning of a protein which inhibits the proliferation of capillary endothelial cells *in vitro* and thus may have potential antiangiogenic effects on solid tumours.

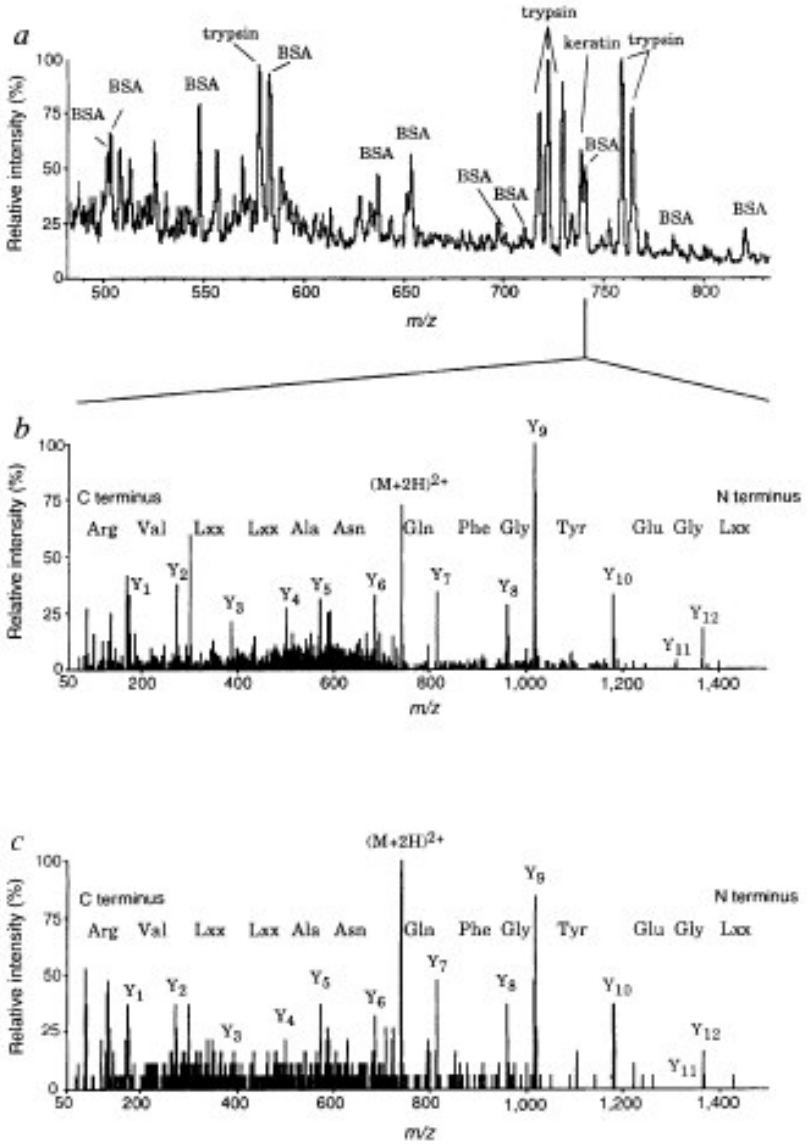


FIG. 1 Testing and sensitivity of the new procedure using BSA. *a*, Peptide ion spectrum of BSA at 800 fmol total material loaded on a gel. Peptide peaks are labelled according to their identification by tandem mass spectrometry. *b*, Fragmentation of the selected doubly charged peptide ion at mass-to-charge *m/z* 740.5. Fragmentation of tryptic peptides at the amide bonds predominantly produces ion series containing the C terminus (designated Y''_1, Y''_2, and so on, see ref. 14). A continuous Y'' ion series[14] could be assigned to the dominant peaks in the spectrum yielding the sequence Lxx-Gly-Glu-Tyr-Gly-Phe-Gln-Asn-Ala-Lxx-Lxx-Val-Arg, where Lxx is either Leu or Ile, which corresponds to a tryptic peptide of BSA. *c*, The same ion as in *b*, obtained in a separate experiment in which 80 fmol BSA was loaded onto a gel that was silver stained.

METHODS. BSA was quantified by amino-acid analysis. Acrylamide gels were prepared using standard protocols and stained with Coomassie blue. *In gel* reduction, acetamidation and tryptic digestion were similar to published procedures[15,16]. After washing with 100 mM NH_4HCO_3 and acetonitrile, gel pieces were swollen in the digestion buffer containing 50 mM NH_4HCO_3, 5 mM $CaCl_2$ and 12.5 ng μl^{-1} trypsin (Boehringer Mannheim, sequencing grade) at 4 °C. After 45 min, the supernatant was aspirated and replaced with 5–10 µl of the same buffer without trypsin to keep gel pieces wet during enzymatic cleavage (37 °C, overnight). Peptides were extracted by three changes of 5% formic acid and acetonitrile and dried down. Approximately 100 nl of POROS R2 sorbent (Perseptive Biosystems) was placed in the tip of a pulled GC 100F-10 (CEI, Pangboume) capillary. Note that the resin is not packed, and that no frit or other micro LC assembling is necessary. A new capillary and a new portion of resin are used for each analysis to avoid cross-contaminations even at the femtomole level. Dried peptide mixture was dissolved in 10 µl 5% formic acid, loaded onto the pre-equilibrated capillary, washed and eluted with 60% methanol in 5% formic acid into the spraying capillary. The elution volume is 10-fold larger than the resin volume, resulting in good peptide recovery. Nano-electrospray was performed on an API III (Perkin-Elmer Sciex, Ontario, Canada) mass spectrometer as described[4,7]. For precursor ion selection, quadrupole 1 was set to transmit a mass window of 2 Da. Step size for the tandem mass spectra was 0.2 Da, and resolution was set so that fragment masses could be assigned to better than 1 Da.

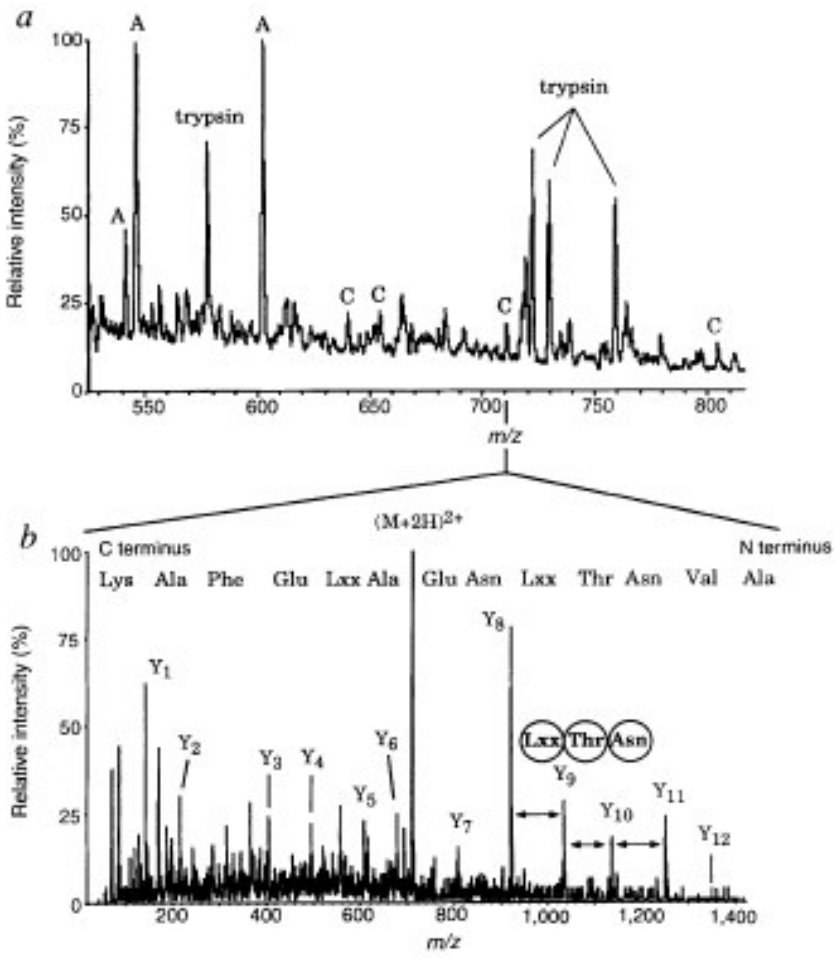


FIG. 2 Identification of protein interaction partners. *a*, Part of the peptide ion spectrum of a tryptic digest of a band of 160K protein excised from polyacrylamide gel. Peptides that were not present in blank were sequenced. The major peaks are tryptic peptides from rabbit antibody light chain (A). *b*, Tandem mass spectrometry of one of the other peptide ions shown in *a*. The masses of the ion and of four fragment ions were assembled by PeptideSearch into a peptide sequence tag[7]. This tag uniquely retrieved carbamoyl-phosphate synthase (*M*, 164K) from a non redundant, inclusive sequence database (NRDB, maintained by C. Sander, EMBL). As shown in *b*, the complete series of Y^{n} ions confirms that the peptide sequence found is correct. Mass spectra of three other ions (labelled C in *a*) identified the same protein.

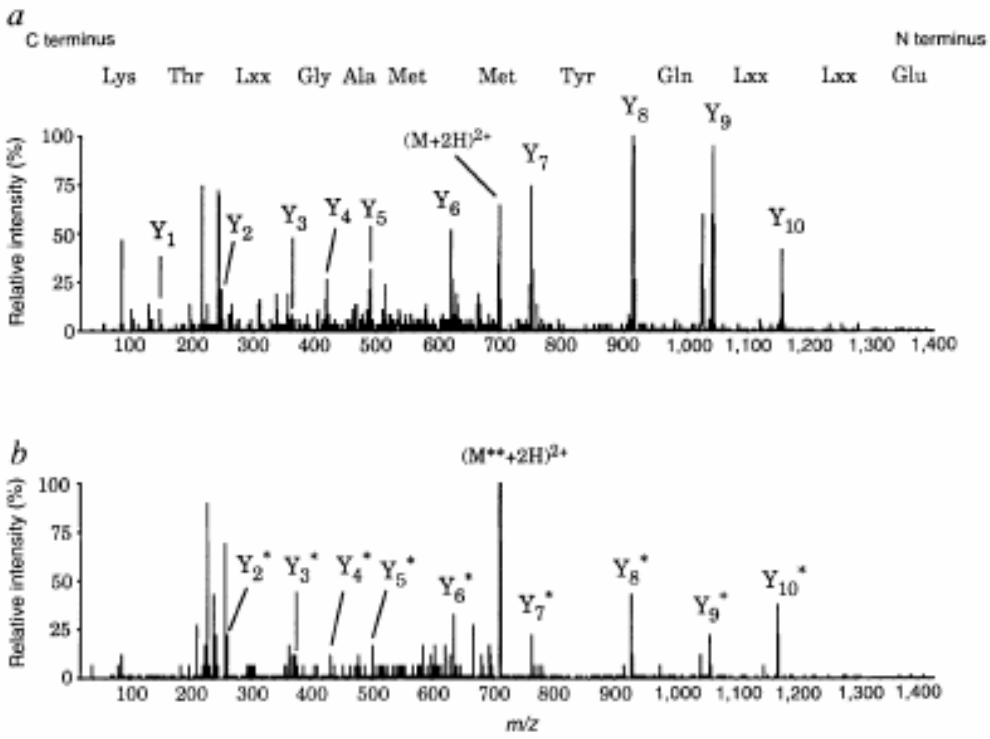


FIG. 3 Sequencing of a peptide from an endothelial cell growth inhibitor. *a*, Tandem mass spectrum of a peptide with *m/z* 699.3 from the tryptic digest. *b*, Tandem mass spectrum of the same peptide sequenced after esterification[6] of the total digest mixture. Esterification results in a characteristic mass shift of the C terminus containing fragment ions (Y" ions) by 14 Da and an additional shift of 14 Da for each Asp or Glu residue. The asterisks indicate the incorporated methyl groups. Software-assisted matching of the spectra in *a* and *b* allowed assignment of the sequence with a high degree of certainty.

METHODS. The protein was purified from 20 l conditioned medium from human neurobastoma cells using the inhibition of proliferation of endothelial cells as an assay. After initial concentration using ultrafiltration, the conditioned medium was subjected to cation-exchanged chromatography and chromatofocusing to enrich the inhibitory activity. Final purification was performed using preparative native PAGE followed by preparative isoelectric focusing in tubes (T.F., in preparation). SDS–PAGE analysis indicated the presence of a protein of *M*, 45K (both under reducing and non-reducing conditions) and a yield of approximately 400 ng. Half of the gel piece containing 200 ng of the purified protein was subjected to the new mass-spectrometric sequencing procedure.

TABLE 1 Comparison of the deduced amino-acid sequence to the partial sequences obtained by tandem mass spectrometry

Nucleic acid sequence of cloned gene	Deduced amino-acid sequence	Amino acid sequence obtained by mass spectrometry
GAA CCT GTT CTT TCA CCT GAA CAC AGA	Glu-Pro-Val-Leu-Ser-Pro-Glu-His-Arg	Glu-Pro-Val-Lxx-Ser-Pro-Glu-His-Arg
GAA TTA ATC CAA TAC ATG ATG GCA GGT ATT ACT AAA	Glu-Leu-Ile-Gln-Tyr-Met-Met-Ala-Gly-Ile-Thr-Lys	Glu-Lxx-Lxx-Gln-Tyr-Met-Met-Ala-Gly-Lxx-Thr-Lys
GAC CCA TTT GCA TCA GTT GGT AAT GGT GTT ACA ATT CAC TAC ATG CGT	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asn-Gly-Val-Thr-Ile-His-Tyr-Met-Arg	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asp-Gly-Val-Thr-Lxx-His-Tyr-Met-Arg
ACT CCA TGA TAC TAT GAC CCA GCA ATG AAA	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys
ACT GAC CTA GAA ACT ATT ACT TTA TTA GCT AAA	Thr-Asp-Leu-Glu-Thr-Ile-Thr-Leu-Leu-Ala-Lys	Thr-Asp-Lxx-Glu-Thr-Lxx-Thr-Lxx-Lxx-Ala-Lys
ATA GTT GCA ATT AAT GTT CCT AAA	Ile-Val-Ala-Ile-Asn-Val-Pro-Lys	Lxx-Val-Ala-Lxx-Asn-Val-Pro-Lys
TGT ATG TCA ATG CCT CTT TCA CGT	Cys-Met-Ser-Met-Pro-Leu-Ser-Arg	Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg

The peptide Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg (Lxx is Leu or Ile), which was at the most C-terminal position (as judged by homology), was used to generate an antisense PCR primer (5′-TGATAAAGGCATTGACATAC-3′). As sense primer, a sequence from the 5′-conserved region of the *Mycoplasma arginini* (5′-AGGAATTCACGTTTATTCAG-3′) was used to obtain a nearly full-length clone using DNA prepared from the contaminated tumour cells as template. The sequence has been deposited with the EMBL data library (accession no. X93471). The only discrepancy observed was the prediction of Asn from the nucleotide sequence of the cloned protein instead of the mass spectrometrically sequenced Asp in one of the peptides (underlined). Because the identity of Asp is absolutely certain owing to the +14 Da mass shift induced on this amino acid by the derivatization procedure in conjunction with the mass-spectrometric data on the underivatized peptide, it is concluded that deamidation of the amino acid must have occurred during the purification procedure, an event not unusual during protein purification. Thus the amino-acid sequence provided by mass spectrometry was entirely correct.

Large-scale analysis of the yeast proteome by multidimensional protein identification technology

Michael P. Washburn[1†], Dirk Wolters[1†], and John R. Yates III[1,2*]

We describe a largely unbiased method for rapid and large-scale proteome analysis by multidimensional liquid chromatography, tandem mass spectrometry, and database searching by the SEQUEST algorithm, named multidimensional protein identification technology (MudPIT). MudPIT was applied to the proteome of the *Saccharomyces cerevisiae* strain BJ5460 grown to mid-log phase and yielded the largest proteome analysis to date. A total of 1,484 proteins were detected and identified. Categorization of these hits demonstrated the ability of this technology to detect and identify proteins rarely seen in proteome analysis, including low-abundance proteins like transcription factors and protein kinases. Furthermore, we identified 131 proteins with three or more predicted transmembrane domains, which allowed us to map the soluble domains of many of the integral membrane proteins. MudPIT is useful for proteome analysis and may be specifically applied to integral membrane proteins to obtain detailed biochemical information on this unwieldy class of proteins.

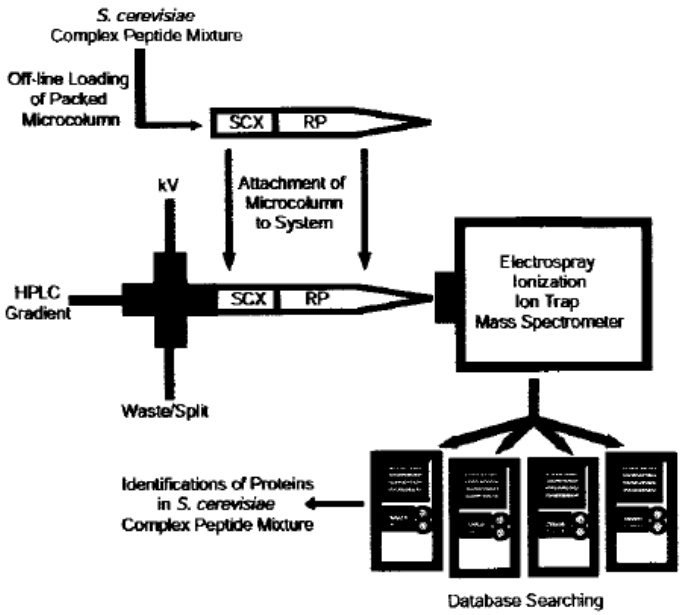


Figure 1. Multidimensional protein identification technology (MudPIT). Based on the method of Link *et al.*[19], complex peptide mixtures from different fractions of a *S. cerevisiae* whole-cell lysate were loaded separately onto a biphasic microcapillary column packed with strong cation exchange (SCX) and reverse-phase (RP) materials. After loading the complex peptide mixture into the microcapillary column, the column was inserted into the instrumental setup. Xcalibur software, HPLC, and mass spectrometer were controlled simultaneously by means of the user interface of the mass spectrometer. Peptides directly eluted into the tandem mass spectrometer because a voltage (kV) supply is directly interfaced with the microcapillary column. As described in the Experimental Protocol, peptides were first displaced from the SCX to the RP by a salt gradient and eluted off the RP into the MS/MS. In an iterative process, the microcolumn was re-equilibrated and an additional salt step of higher concentration displaced peptides from the SCX to the RP. Peptides were again eluted by an RP gradient into the MS/MS, and the process was repeated. The tandem mass spectra generated were correlated to theoretical mass spectra generated from protein or DNA databases by the SEQUEST algorithm[21].

MudPIT

- 100 μm i.d. fuse silica

Link et al, Nat Biotech (1999)
Washburn, Wolters, Yates, Nat Biotech (2001)

MudPIT

- 100 µm i.d. fuse silica
- Tip is pulled to ~5 µm I.D.

MudPIT

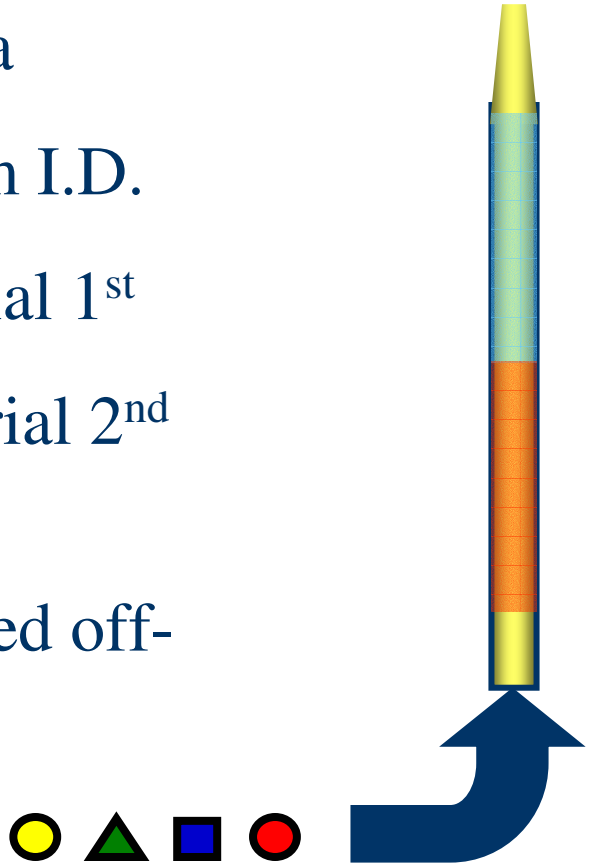
- 100 µm i.d. fuse silica
- Tip is pulled to ~5 µm I.D.
- Pack with **C18** material 1st

MudPIT

- 100 µm i.d. fuse silica
- Tip is pulled to ~5 µm I.D.
- Pack with **C18** material 1st
- Pack with **SCX** material 2nd

MudPIT

- 100 µm i.d. fuse silica
- Tip is pulled to ~5 µm I.D.
- Pack with **C18** material 1st
- Pack with **SCX** material 2nd
- Peptide digest is loaded off-line



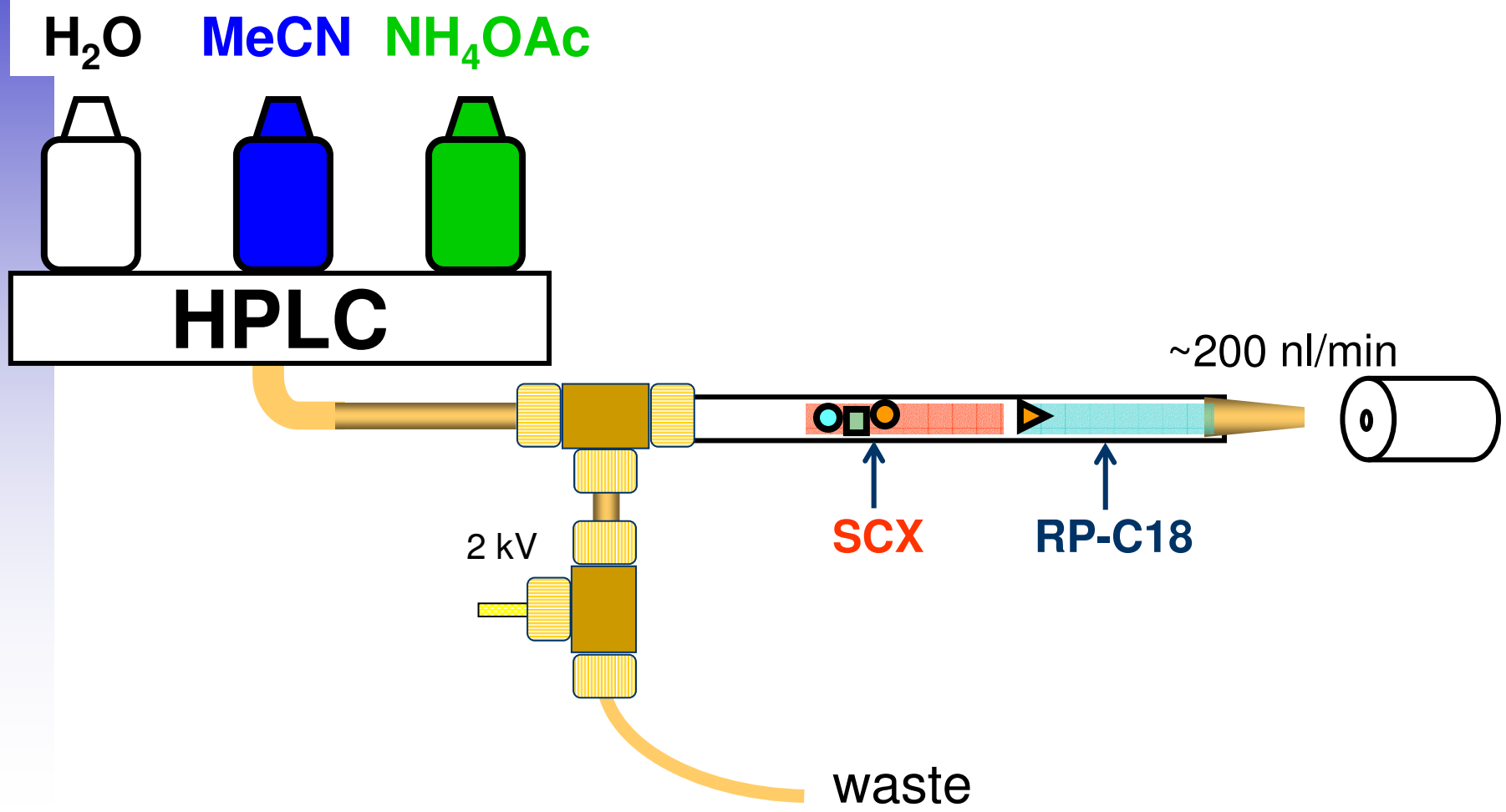
MudPIT

- 100 µm i.d. fuse silica
- Tip is pulled to ~5 µm I.D.
- Pack with **C18** material 1st
- Pack with **SCX** material 2nd
- Peptide digest is loaded off-line

Multidimensional Protein Identification Technology (MudPIT)

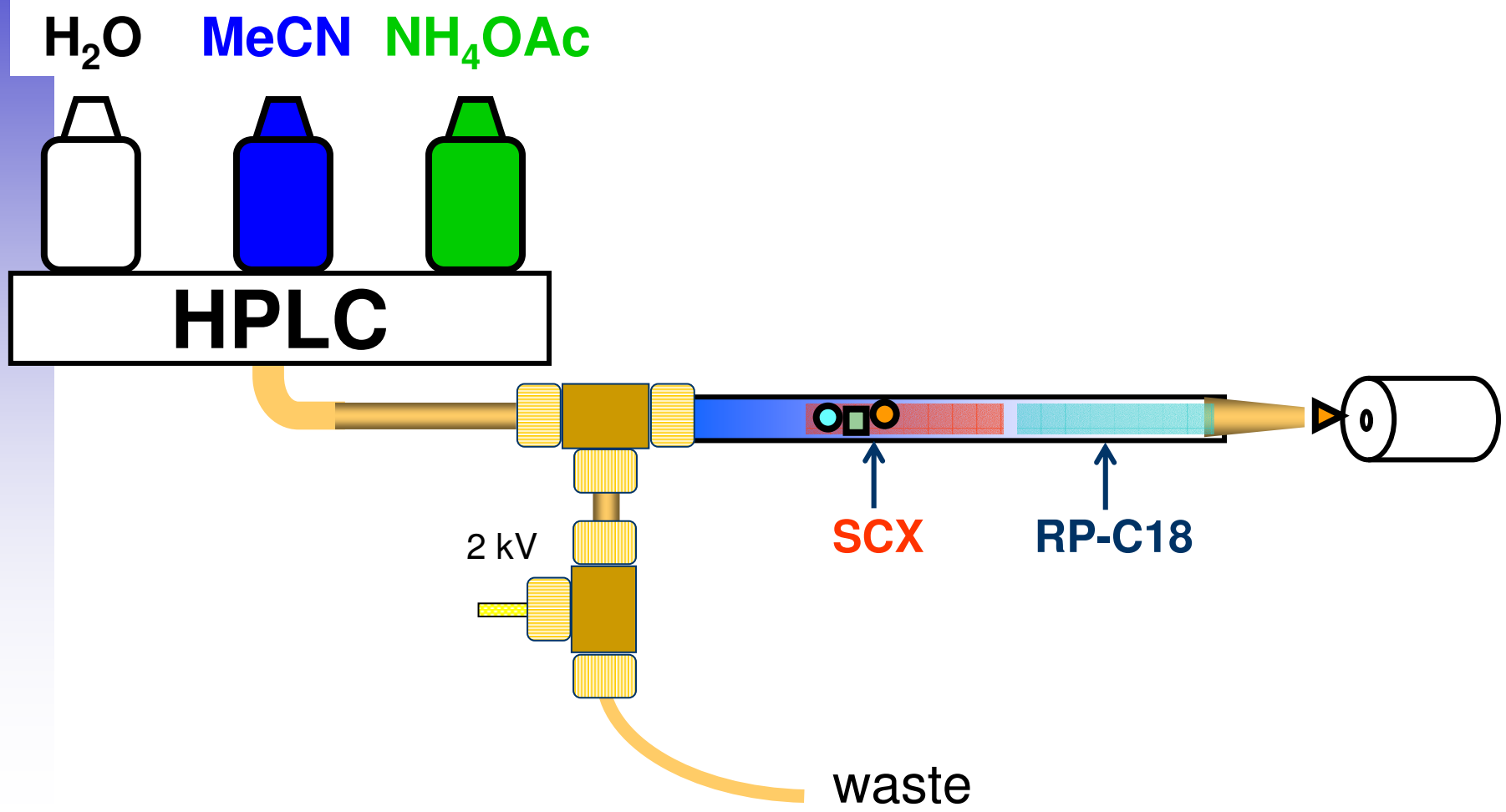
MudPIT Column

MudPIT



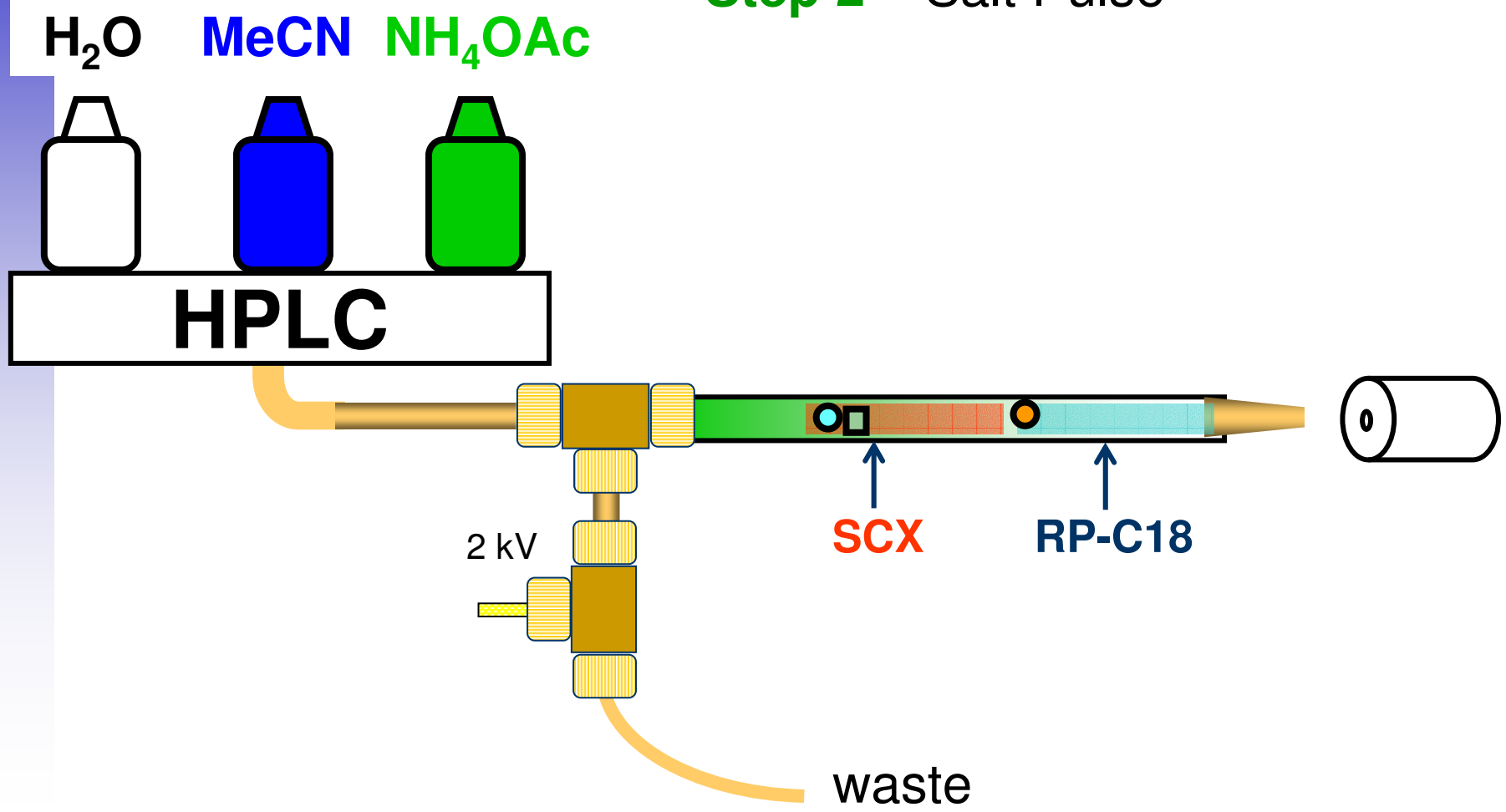
MudPIT

- **Step 1** – Reverse Phase Gradient



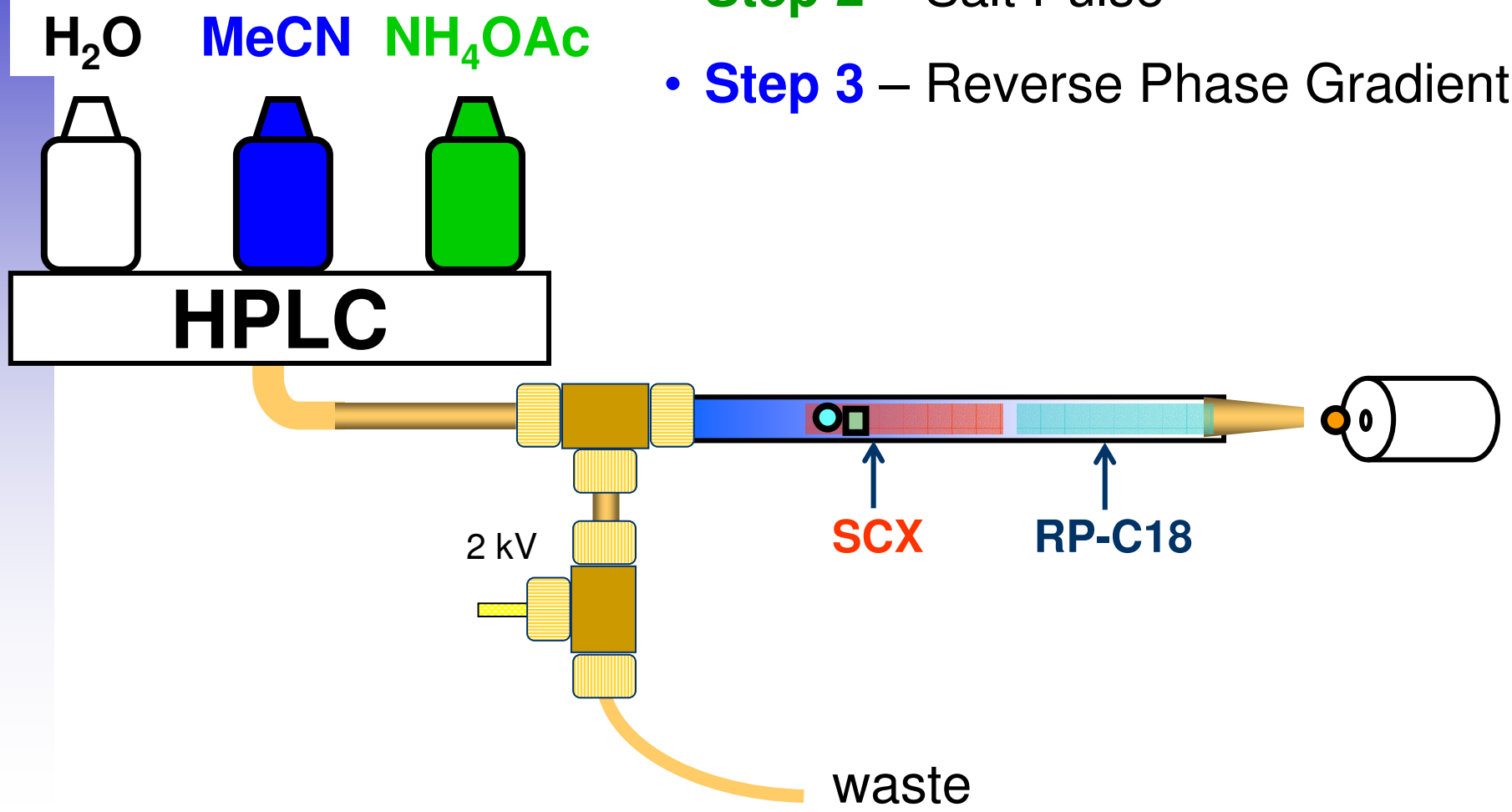
MudPIT

- **Step 1** – Reverse Phase Gradient
- **Step 2** – Salt Pulse



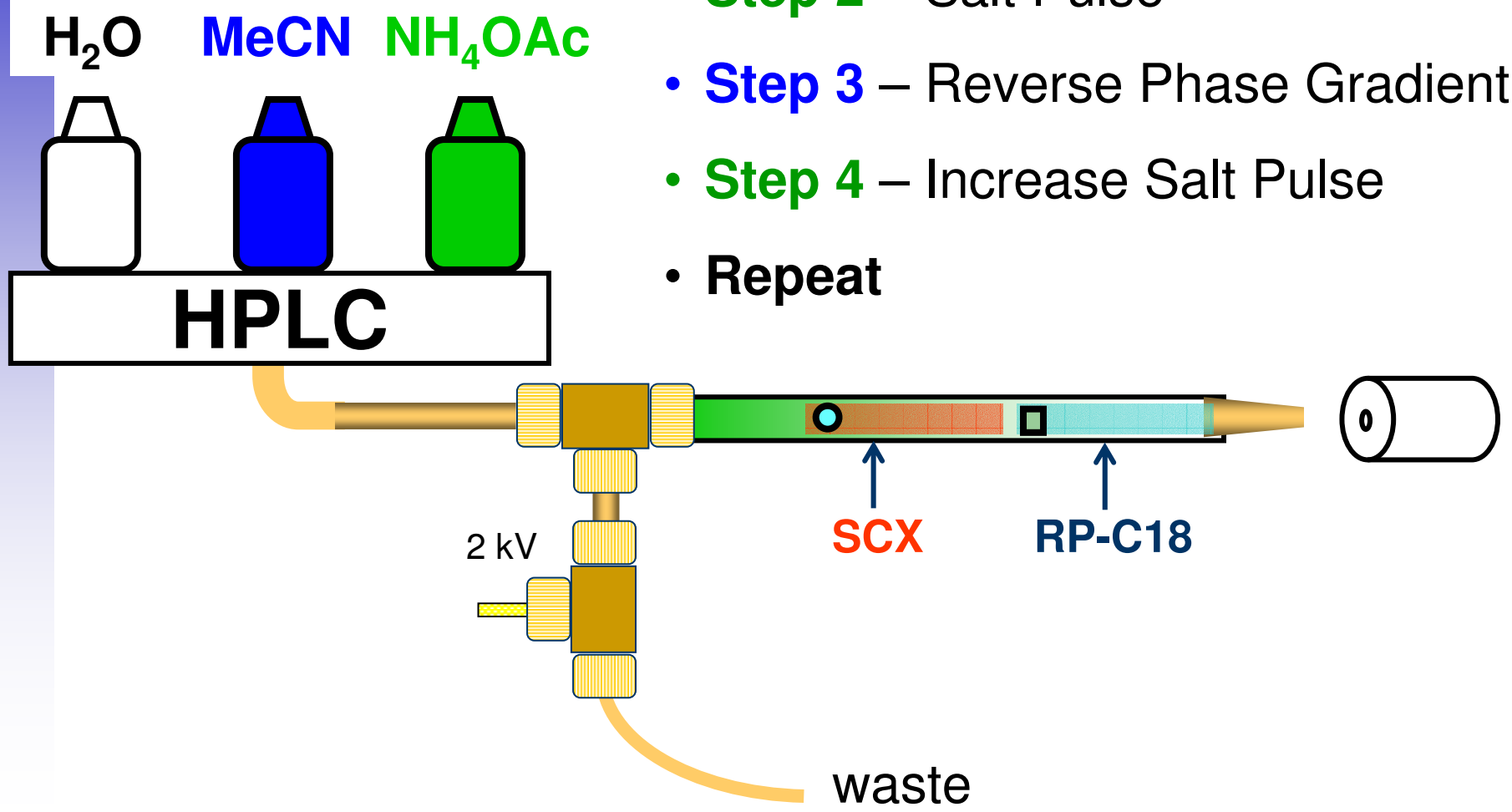
MudPIT

- **Step 1** – Reverse Phase Gradient
- **Step 2** – Salt Pulse
- **Step 3** – Reverse Phase Gradient

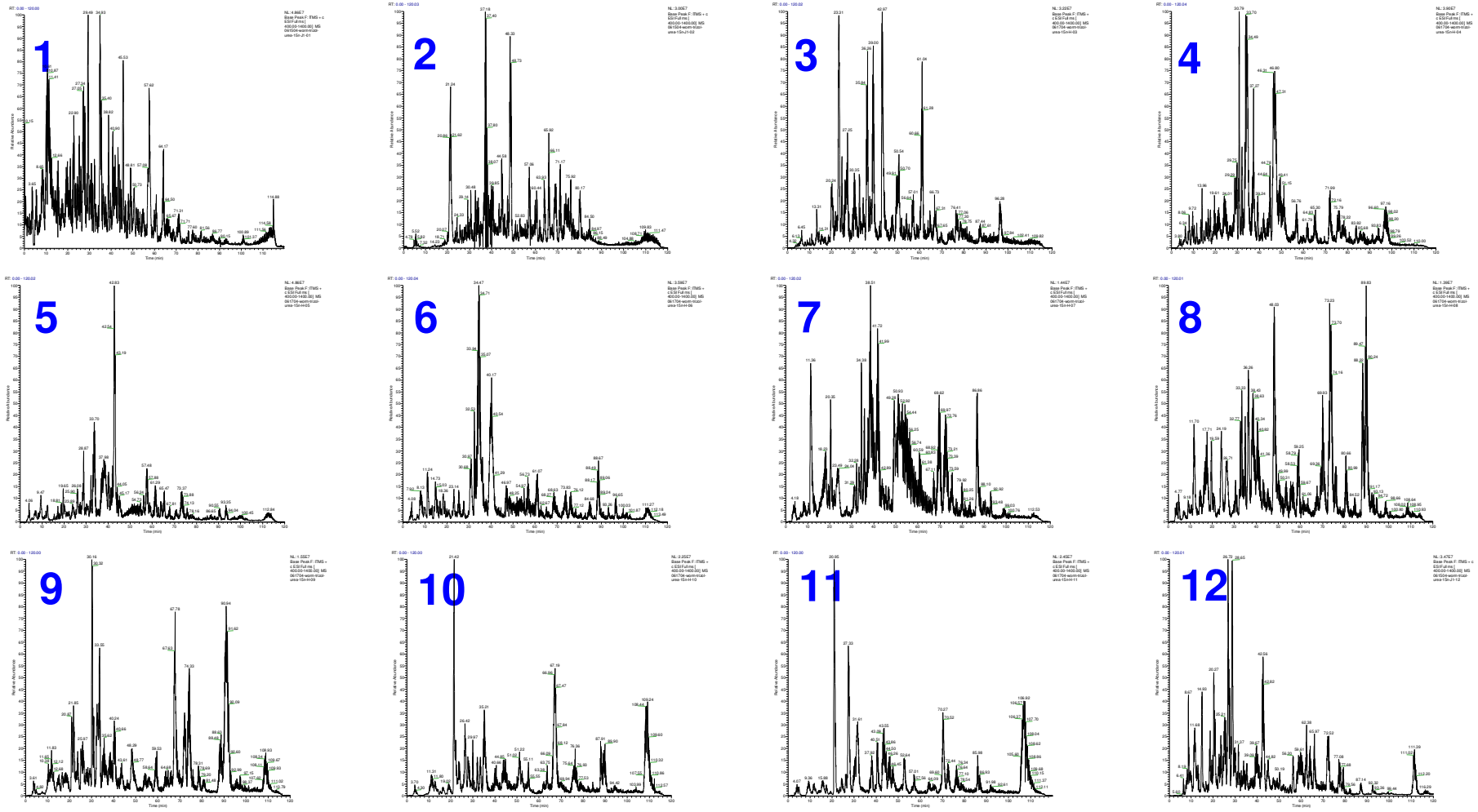


MudPIT

- **Step 1** – Reverse Phase Gradient
- **Step 2** – Salt Pulse
- **Step 3** – Reverse Phase Gradient
- **Step 4** – Increase Salt Pulse
- **Repeat**



µLC/µLC/MS/MS of *C. elegans* Peptides



MS/MS Spectra: 280290

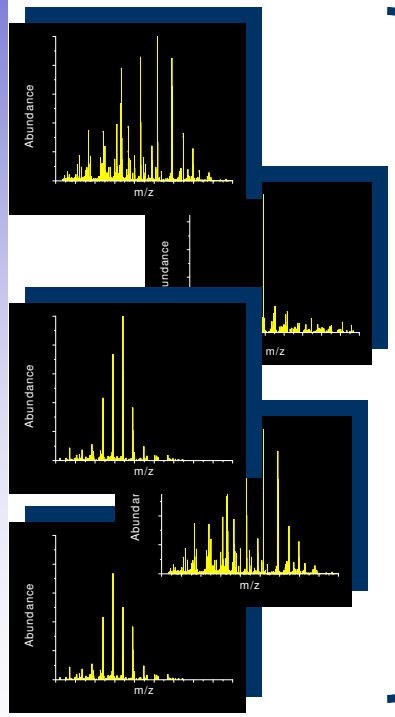
Peptides Identified: 2753

Proteins Identified

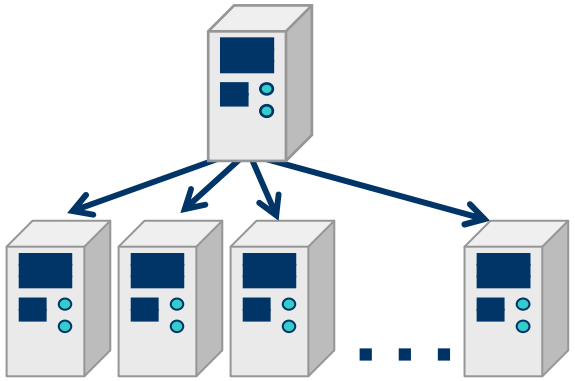
With >95% Confidence: 941

Qualitative Analysis of Tandem Mass Spectra

100,000's of MS/MS Spectra



Database Search Using SEQUEST



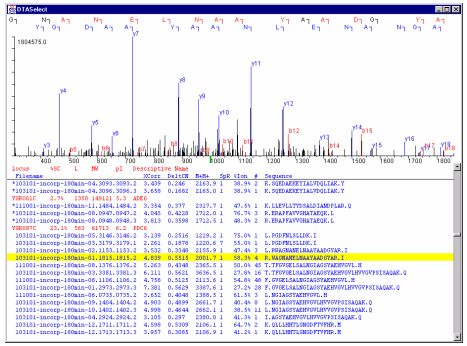
Beowulf Computer Cluster

Peptide IDs

Spectrum 1: ERTYILDFGH
Spectrum 2: WEQTMNVCSW
Spectrum 3: ERTYEWIPWQ
Spectrum 4: …

Output Sorting and Filtering

Protein Identifications



Assembly Using DTASelect

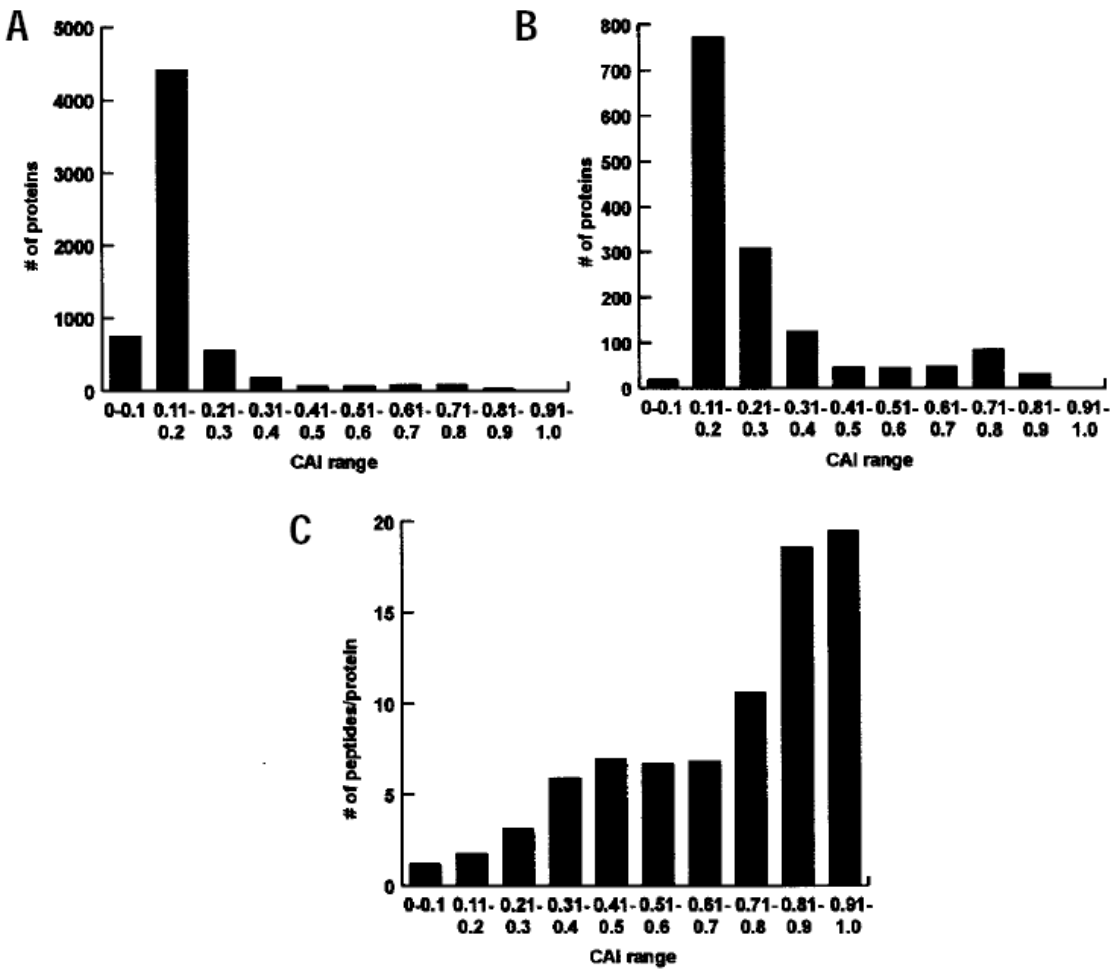


Figure 2. Codon adaptation index (CAI) distribution of the identified *S. cerevisiae* proteome and the predicted *S. cerevisiae* genome. (A) CAI distribution of the proteins predicted in the *S. cerevisiae* genome. (B) Compare this to the distribution of the proteins identified in this study over CAI ranges. In both cases, the largest protein region is found between the CAI range of 0.11 and 0.2. (C) The average number of peptides identified for each protein in a particular CAI range was determined and plotted against CAI ranges.

Table 1. Known subcellular localization of proteins identified in *S. cerevisiae* fractions[a]

Subcellular compartment	Soluble fraction[b]	Lightly washed insoluble fraction[b]	Heavily washed insoluble fraction[b]
Cell wall	2	1	1
Plasma membrane	5	18	35
Cytoplasm	286	264	274
Cytoskeleton	11	20	22
Endoplasmic reticulum	12	36	42
Golgi	3	10	16
Transport vesicles	4	14	16
Nucleus	67	122	151
Mitochondria	43	87	83
Peroxisome	2	3	3
Endosome	1	1	2
Vacuole	5	10	6
Microsomes	0	0	1
Lipid particles	0	2	3

[a]Subcellular localizations obtained from the *S. cerevisiae* subcellular localization catalog at the Munich Information Center for Protein Sequences website [24].
[b]Proteins identified in individual runs were analyzed for their subcellular localization. The subcellular localization of many of the proteins detected and identified is unknown. Therefore, not all of the proteins detected and identified are represented in this table.

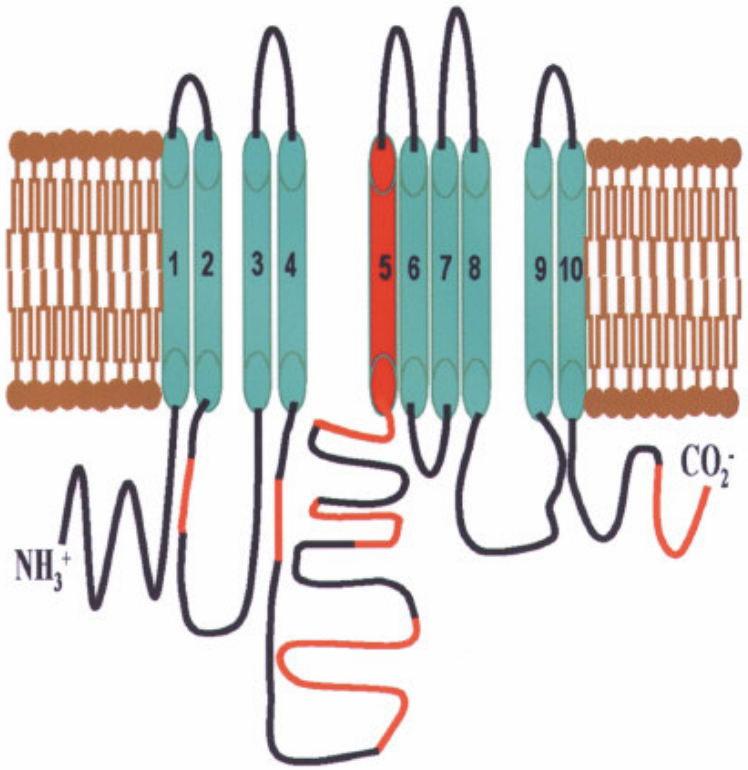


Figure 3. Peptide mapping of the integral membrane protein PMA1. A two-dimensional representation of PMA1 is displayed. Cylinders represent the predicted Tm domains as reported by MIPS (ref. 24). The protein segments between predicted Tm domains are drawn to approximate scale. Black lines and green cylinders represent segments of the protein not identified in this study. Red lines and the red cylinder represent segments of the protein identified in this study. One peptide was detected and identified between Tm domains 2 and 3, 10 peptides were detected and identified between Tm domains 4 and 5, and one peptide was detected and identified in the C terminus. We also detected and identified a peptide corresponding to Tm domain 5 in our analysis. The 320-amino acid domain between Tm domains 4 and 5 is the largest in the protein.

Table 2. Proteins identified containing three or more predicted transmembrane domains[a]

Number of predicted transmembrane domains	Number of proteins in class	Number of proteins in class identified by MudPIT	Percentage of total predicted
3	185	31	17
4	101	16	16
5	57	12	21
6	58	14	24
7	56	7	13
8	54	13	24
9	71	12	17
10	53	14	26
11	30	4	13
12	15	4	27
13	8	3	38
14	3	0	0
15	4	1	25
16	1	0	0
20	1	0	0
Totals	697	131	19

[a]The Munich Information Center for Protein Sequences website was used to obtain this information[24]. The prediction of transmembrane domains at this site is based on Klein *et al.*[34] and Goffeau *et al*[35].

Table 3. Proteins identified with 10 or more predicted transmembrane (Tm) domains[a]

Locus	Name	No. of predicted Tm domains	No. of peptides identified	Peptide hits within Tm domains[b]	Peptide hits to largest soluble domain	CAI	MW (kDa)	Membrane localization in cell
YCR017C	–	15	2	N	1	0.16	108	–
YGR032w	GSC2	13	4	N	3	0.21	217	Plasma
YIL030c	SSM4	13	1	N	0	0.17	151	–
YJL198w	–	13	1	N	1	0.18	98	–
YDR135c	YCF1	12	3	N	2	0.15	171	Vacuolar
YKL209c	STE6	12	1	N	1	0.13	145	Plasma
YLL015w	–	12	1	N	0	0.14	177	–
YLR342w	FKS1	12	6	1 P	3	0.27	215	Plasma
YGL022w	STT3	11	2	N	2	0.21	82	ER[c]
YNL268w	LYP1	11	1	N	1	0.22	68	Plasma
YNR013c	–	11	3	1 P	1	0.19	99	Plasma
YPL058c	PDR12	11	6	N	3	0.29	171	–
YBR068c	BAP2	10	2	N	0	0.16	68	Plasma
YBR243c	ALG7	10	2	1 P, 1 C	0	0.13	50	ER
YDR342c	HXT7	10	1	N	1	0.52	63	Plasma
YDR343c	HXT6	10	2	N	1	0.52	63	Plasma
YDR345c	HXT3	10	1	N	1	0.49	63	Plasma
YDR497c	ITR1	10	1	C	0	0.19	64	Plasma
YER119c	–	10	1	P	0	0.10	49	–
YFL025c	BST1	10	1	N	0	0.13	118	ER
YGL008c	PMA1	10	13	1 C	10	0.73	100	Plasma
YGR125w	–	10	1	1 P	0	0.12	117	–
YHR094c	HXT1	10	1	N	1	0.41	63	Plasma
YLL061w	MMP1	10	1	N	0	0.13	64	–
YOR328w	PDR10	10	1	N	1	0.13	176	Plasma
YPL036w	PMA2	10	11	1 C	10	0.30	102	Plasma

[a]The Munich Information Center for Protein Sequences website was used to obtain this information. The prediction of transmembrane domains at this site is based on Klein *et al.*[34] and Goffeau *et al.*[35] [b]Abbreviations: N, none; P, partially covers a transmembrane domain; C, completely covers a transmembrane domain. [c]Endoplasmic reticulum.

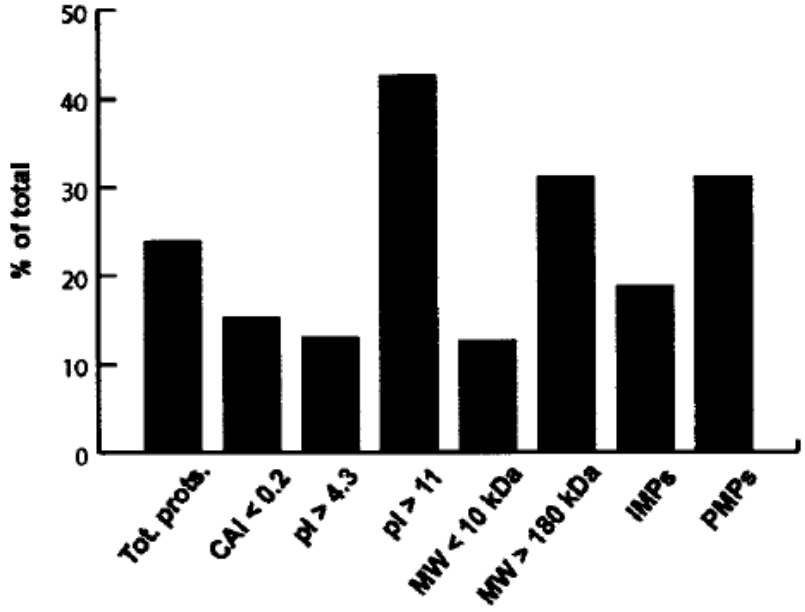


Figure 4. Sensitivity of MudPIT to a wide variety of protein classes. The percentage of proteins identified in this study from a variety of protein classes is presented. The percentages were determined by dividing the number of proteins identified in the study in each category shown by the total number of predicted proteins from each category shown. MIPS (ref. 24) and the Yeast Proteome Database[33] were used to obtain the predicted numbers of proteins from *S. cerevisiae* in each class. From left to right are the percentages identified of total proteins, proteins with a CAI <0.2, proteins with a pI < 4.3, proteins with a pI >11, proteins with a MW <10kDa, proteins with a MW >180 kDa, integral membrane proteins (IMPs) with three or more predicted transmembrane domains, and peripheral membrane proteins (PMPs).

Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins*

Leigh Anderson‡§ and Christie L. Hunter¶

Quantitative LC-MS/MS assays were designed for tryptic peptides representing 53 high and medium abundance proteins in human plasma using a multiplexed multiple reaction monitoring (MRM) approach. Of these, 47 produced acceptable quantitative data, demonstrating within-run coefficients of variation (CVs) (n = 10) of 2–22% (78% of assays had CV <10%). A number of peptides gave CVs in the range 2–7% in five experiments (10 replicate runs each) continuously measuring 137 MRMs, demonstrating the precision achievable in complex digests. Depletion of six high abundance proteins by immunosubtraction significantly improved CVs compared with whole plasma, but analytes could be detected in both sample types. Replicate digest and depletion/digest runs yielded correlation coefficients (R^2) of 0.995 and 0.989, respectively. Absolute analyte specificity for each peptide was demonstrated using MRM-triggered MS/MS scans. Reliable detection of L-selectin (measured at 0.67 μg/ml) indicates that proteins down to the μg/ml level can be quantitated in plasma with minimal sample preparation, yielding a dynamic range of 4.5 orders of magnitude in a single experiment. Peptide MRM measurements in plasma digests thus provide a rapid and specific assay platform for biomarker validation, one that can be extended to lower abundance proteins by enrichment of specific target peptides (stable isotope standards and capture by anti-peptide antibodies (SISCAPA)). *Molecular & Cellular Proteomics 5:573–588, 2006.*

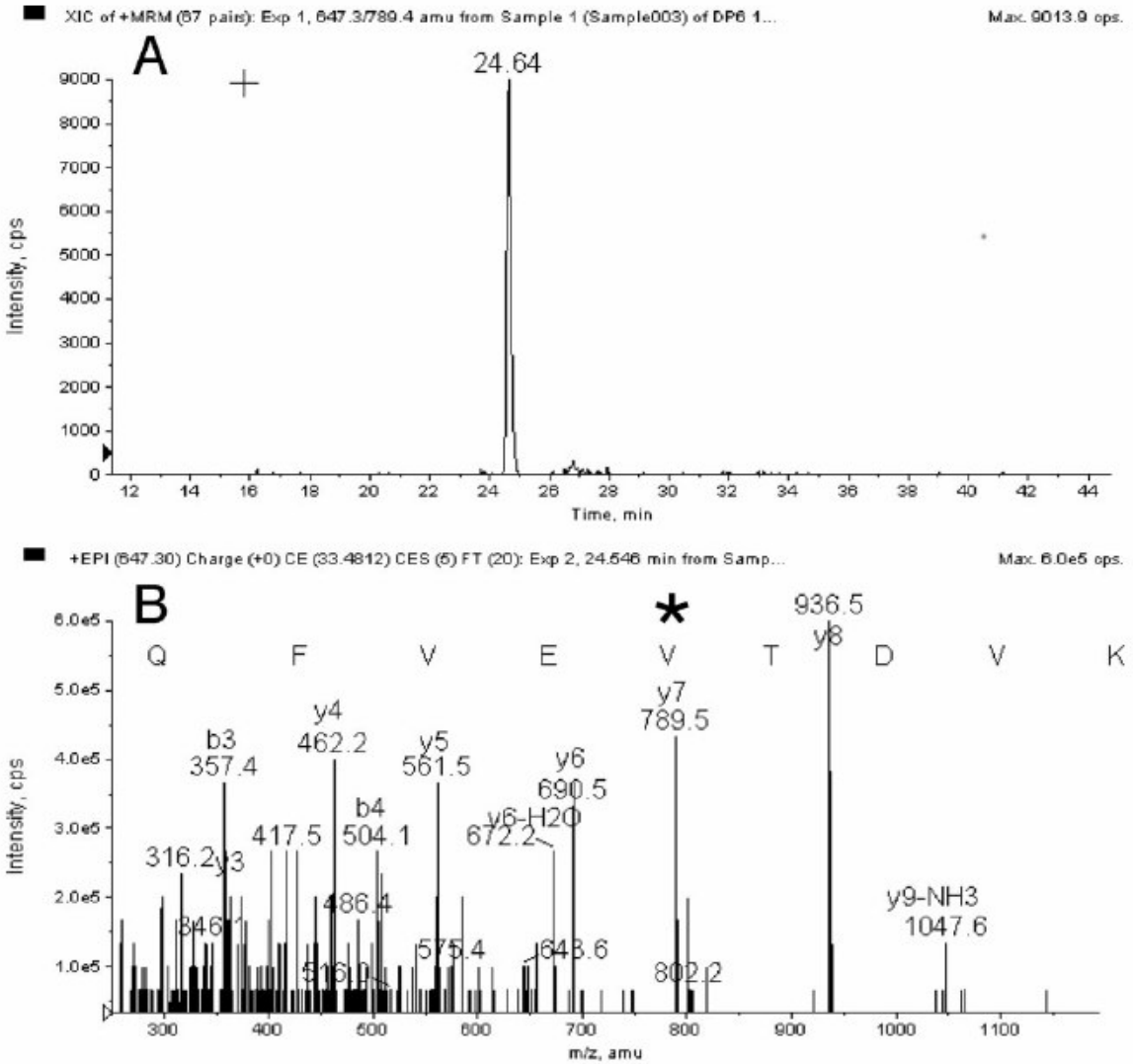


FIG. 1. **MIDAS workflow (MRM-triggered MS/MS) verification of the identity of peptide DLQFVEVTDVK representing fibronectin (normal concentration, ~300 μg/ml) in a digest of depleted human plasma.** *A* shows the ion current profile of MRM 647.3/789.4 (MS1/MS2), and *B* shows the MS/MS spectrum of peptide fragments taken at the time of the peak (* marks the y7 fragment monitored in this MRM). *cps*, counts per second.

TABLE I

Summary design of data sets (experiments A–F)

Replicate runs were performed in series with 30-min washes between runs. Load is expressed as the equivalent volume of plasma from which the sample was derived. Load factors express total or non-depleted (MARS flow-through) loads relative to experiment A.

Experiment	Replicate runs	Sample	LC system	Equivalent plasma volume loaded	Load factor		PolySIS spike
					Total protein	Non-depleted proteins	
				μl			*fmol*
A	10	Depleted plasma digest	LC Packings	0.01	1	1	1.3
B	10	Whole plasma digest	LC Packings	0.01	10	1	1.3
C	10	Whole plasma digest	Eksigent	0.001	6	0.1	2.0
D	10	Depleted plasma digest	Eksigent	0.01	0.6	1	2.0
E	10	Depleted plasma digest	Eksigent	0.033	3.3	3.3	6.0
F1_1	4	Depletion 1, digest 1	Eksigent	0.01	1	1	
F1_2	4	Depletion 1, digest 2	Eksigent	0.01	1	1	
F2_1	4	Depletion 2, digest 1	Eksigent	0.01	1	1	
F2_2	4	Depletion 2, digest 2	Eksigent	0.01	1	1	

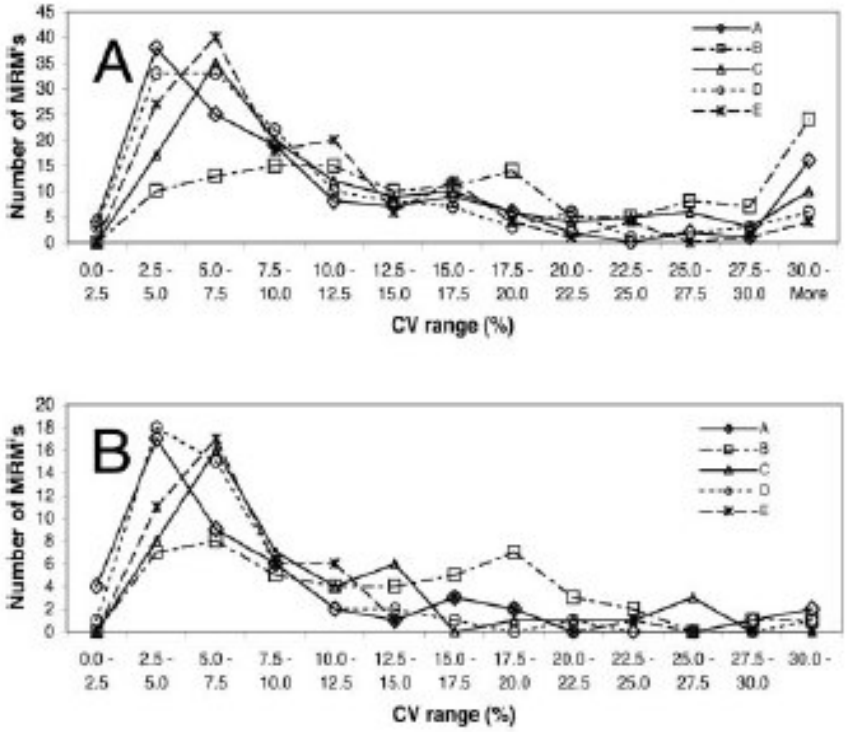


FIG. 2. **Histogram of CVs of MRM values (peak areas) for five experiments (experiments A–E) across all 137 MRMs (*A*) or across the 47 "best" MRMs (*B*).**

TABLE II

A set of MRMs designed and tested for the detection of 53 proteins in human plasma or serum

Protein name, tryptic peptide sequence, retention time in experiment D, peptide mass (MS1, *m/z*), singly charged fragment MS2 (*m/z*), mean peak area values, and CVs for 10 replicate analyses across five experiments are shown.

MRM Count	Protein	Peptide Sequence	RT in D	Best MRM	SIS	MS1/MS2	Mean Peak Areas					CV (%)				
							A	B	C	D	E	A	B	C	D	E
1	Afamin	DADPDTFFAK	21.5	X		563.8 / 825.4	1.7E+05	2.0E+05	3.2E+04	1.6E+05	3.7E+05	8	9	10	4	6
2						563.8 / 940.4	3.8E+04	5.9E+03	3.5E+04	3.1E+04	7.0E+04	7	15	17	10	10
3	Alpha-1-acid glycoprotein 1	NWGLSVYADKPETTK	19.7			570.3 / 1052.5	2.3E+05	5.9E+04	1.9E+05	1.6E+05	3.1E+05	6	12	9	8	17
4				X		570.3 / 575.3	3.9E+05	9.7E+04	3.3E+05	3.0E+05	5.9E+05	3	13	7	8	16
5					X	575.6 / 1068.5	1.4E+04	2.3E+04	6.4E+03	9.0E+03	2.5E+04	11	21	11	25	16
6	Alpha-1-antichymotrypsin	EIGELYLPK	22.4			531.3 / 633.4	5.0E+05	1.2E+05	7.5E+05	4.3E+05	9.9E+05	5	6	6	3	5
7				X		531.3 / 819.5	7.5E+05	1.6E+05	1.1E+06	5.7E+05	1.3E+06	2	5	7	2	5
8					X	535.3 / 827.5	4.7E+04	1.6E+05	5.7E+04	8.3E+04	2.0E+05	5	12	3	3	6
9	Alpha-1B-glycoprotein	LETPDFQLFK	27.7			619.4 / 995.5	2.1E+05	2.7E+04	1.9E+05	1.2E+05	3.0E+05	4	7	11	7	14
10				X		619.4 / 894.5	5.2E+05	7.3E+04	5.0E+05	3.0E+05	8.1E+05	3	7	7	6	11
11	Alpha-2-antiplasmin	LGNQEPGGQTALK	12.6	X		656.8 / 771.4	3.5E+05	7.8E+04	1.5E+05	2.2E+05	5.7E+05	4	20	10	6	8
12						656.8 / 900.5	3.7E+04	7.1E+03	1.4E+04	1.9E+04	4.9E+04	9	38	19	15	10
13					X	660.8 / 779.4	1.8E+05	5.1E+05	4.4E+04	3.1E+05	9.5E+05	6	20	7	7	8
14	alpha-1-antitrypsin	DTEEEDFHVDQVTTVK	17.4			631.3 / 790.4	8.5E+03	2.9E+03	3.5E+03	3.0E+03	1.8E+04	20	33	36	34	23
15						631.3 / 889.5	1.3E+04	3.5E+03	4.5E+03	6.6E+03	4.1E+04	11	33	33	29	13
16	alpha-2-macroglobulin	LLIYAVLPTGDVIGDSAK	36.5	X		923.0 / 1059.5	4.1E+05	5.3E+04	1.6E+04	4.2E+05	4.3E+05	17	20	19	6	8
17						923.0 / 1172.6	1.5E+05	1.9E+04	6.2E+03	1.4E+05	1.4E+05	16	31	20	10	8
18	Angiotensinogen	ALQDQLVLVAAK	23.5			634.9 / 956.6	3.3E+04	1.6E+04	2.2E+03	1.6E+04	1.8E+04	10	56	8	8	13
19						634.9 / 713.5	3.8E+04	2.0E+04	3.0E+03	2.4E+04	2.3E+04	7	51	16	10	16
20					X	638.9 / 964.6	2.7E+04	7.3E+04	1.3E+05	4.5E+04	5.4E+04	9	17	7	10	8
21		PKDPTFIPAPIQAK	19.1			508.3 / 724.4	2.8E+04	3.3E+04	5.0E+04	1.6E+04	2.9E+04	14	17	8	15	11
22				X		508.3 / 556.4	3.0E+04	1.2E+05	5.7E+04	1.8E+04	3.6E+04	10	19	6	14	15

TABLE III

Improvement of CV obtained by combining peak areas of two fragment MRMs versus keeping them separate

Avg, average; frag(s), fragment(s).

Experiment	Avg CV		
	Sum of frags	Frag 1	Frag 2
		%	
A	10.5	11.8	14.8
B	16.2	20.0	19.4
C	11.0	13.0	14.4
D	8.0	9.4	12.3
E	8.5	9.4	11.9

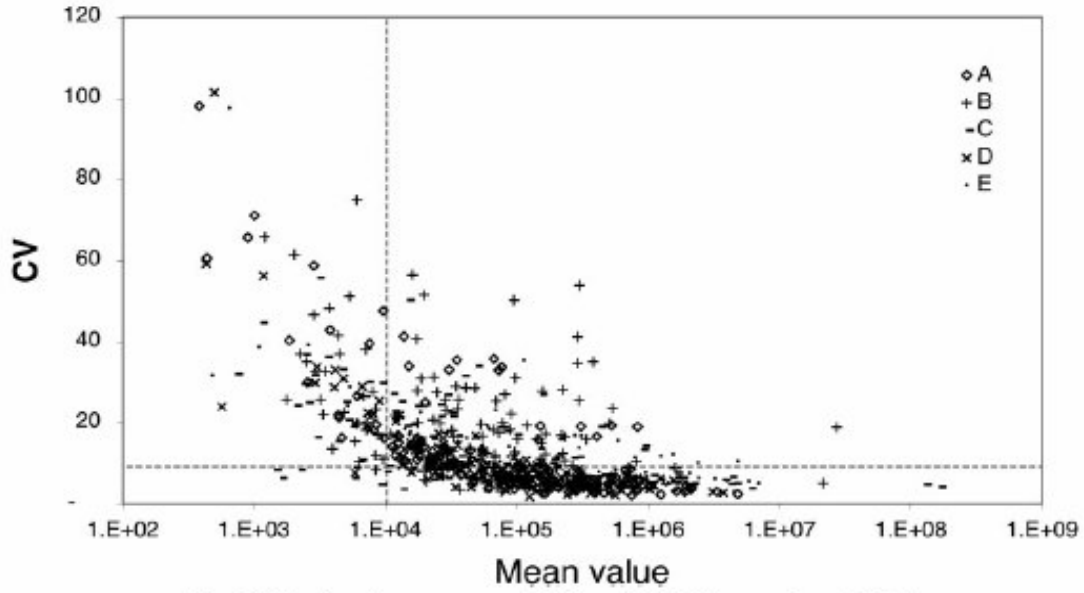
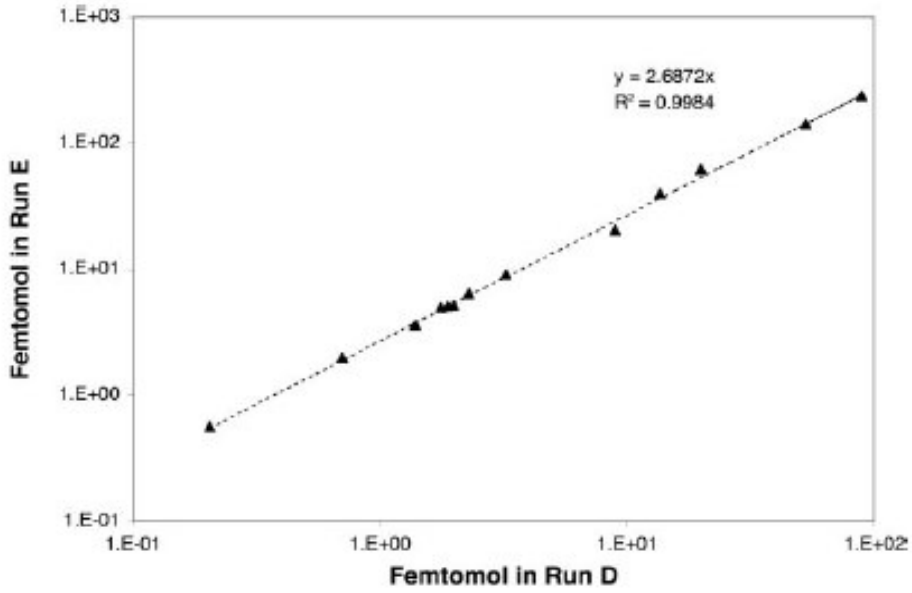


FIG. 3. **Plot of peak area *versus* CV for all MRMs in experiments A–E.**

FIG. 4. **Comparison of computed amounts (based on ratios between Nat and SIS peptides) for 13 peptides in experiments D and E.** Data are included for those ratios where both numerator and denominator peak areas were >1E+04 in experiment D (13 peptides).



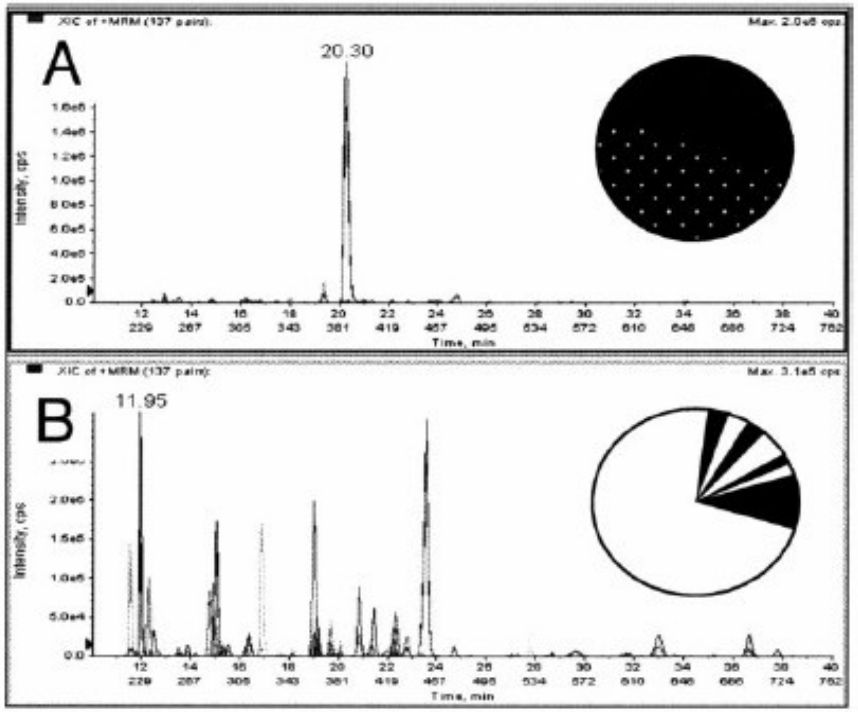


FIG. 5. **Total ion current profiles across the chromatographic peptide separation for digests of undepleted (*A*, whole) and depleted (*B*) plasma from experiments C and D, respectively.** Pie charts represent the protein composition of the samples: whole plasma contains >50% albumin (*stippled region*), whereas the proteins remaining after MARS depletion include fibrinogen, α_2-macroglobulin, complement C3, and all other lower abundance proteins (four remaining pie slices ordered *clockwise* from 12 o'clock), showing the proportion of protein removed by depletion (*white segments* in *B*). The major peptide in *A* at 20.3 min is derived from albumin (whose abundance is shown by the *black* pie segment with *white dots*). *cps*, counts per second.

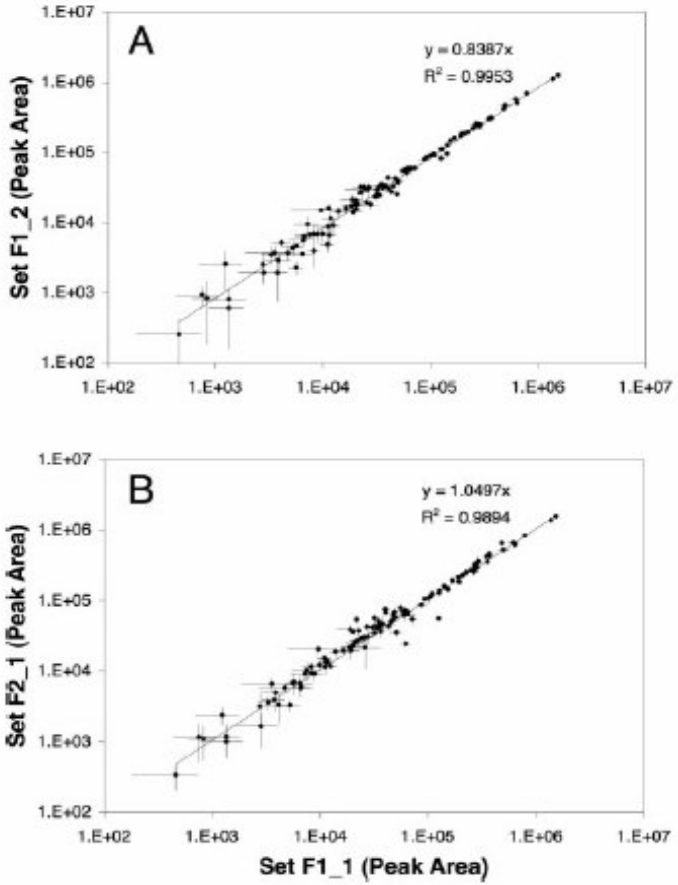


FIG. 6. **Comparison of mean peak areas values for all MRMs in two replicate digests (*A*) and two depletion runs (and subsequent digests) (*B*).** *Error bars* show 1 standard deviation computed from replicate runs.

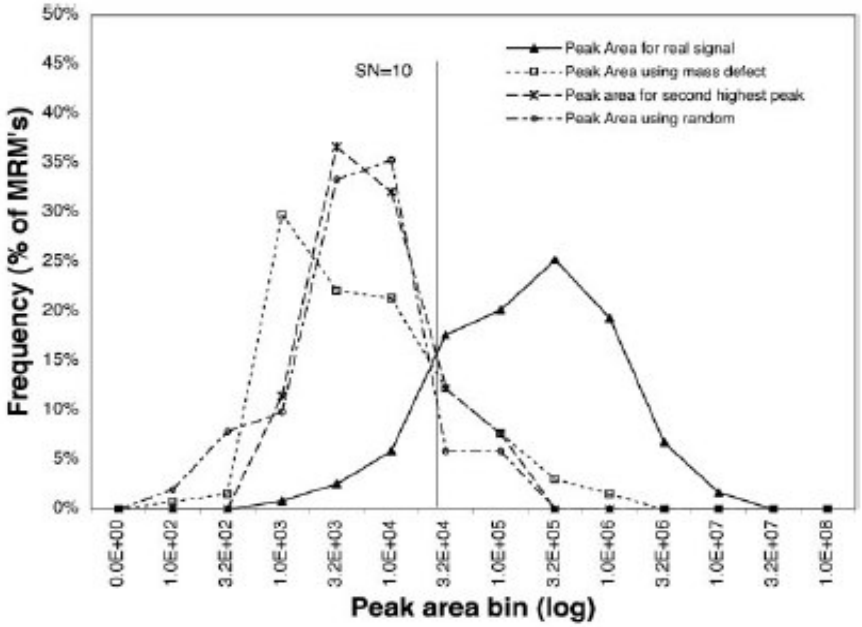


FIG. 7. **Histograms of peak areas obtained with four sets of MRMs in a digest of depleted plasma.**

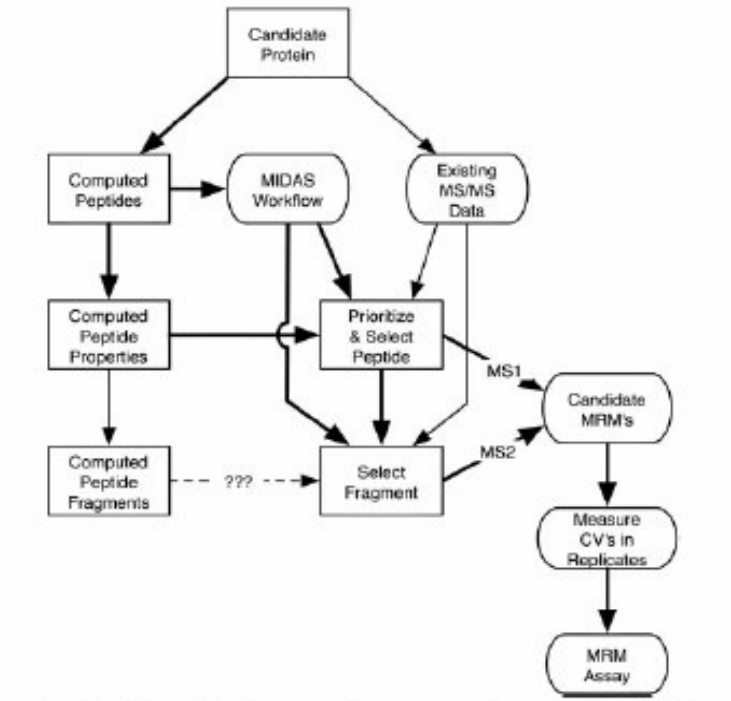


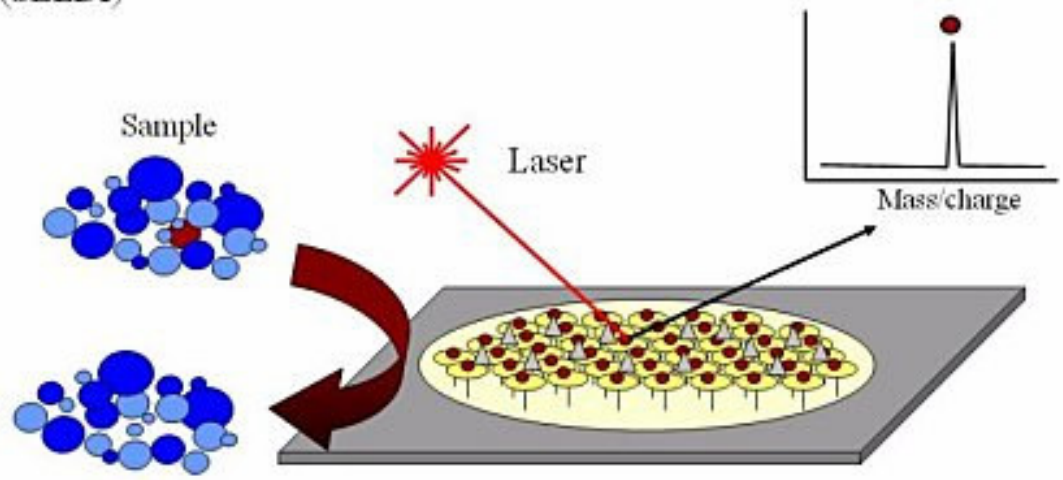
FIG. 8. **Schematic diagram of a process for selecting MRMs.** *Thicker arrows* represent the preferred process. *Square-cornered boxes* represent informatics components, and *rounded-corner boxes* are experimental steps.

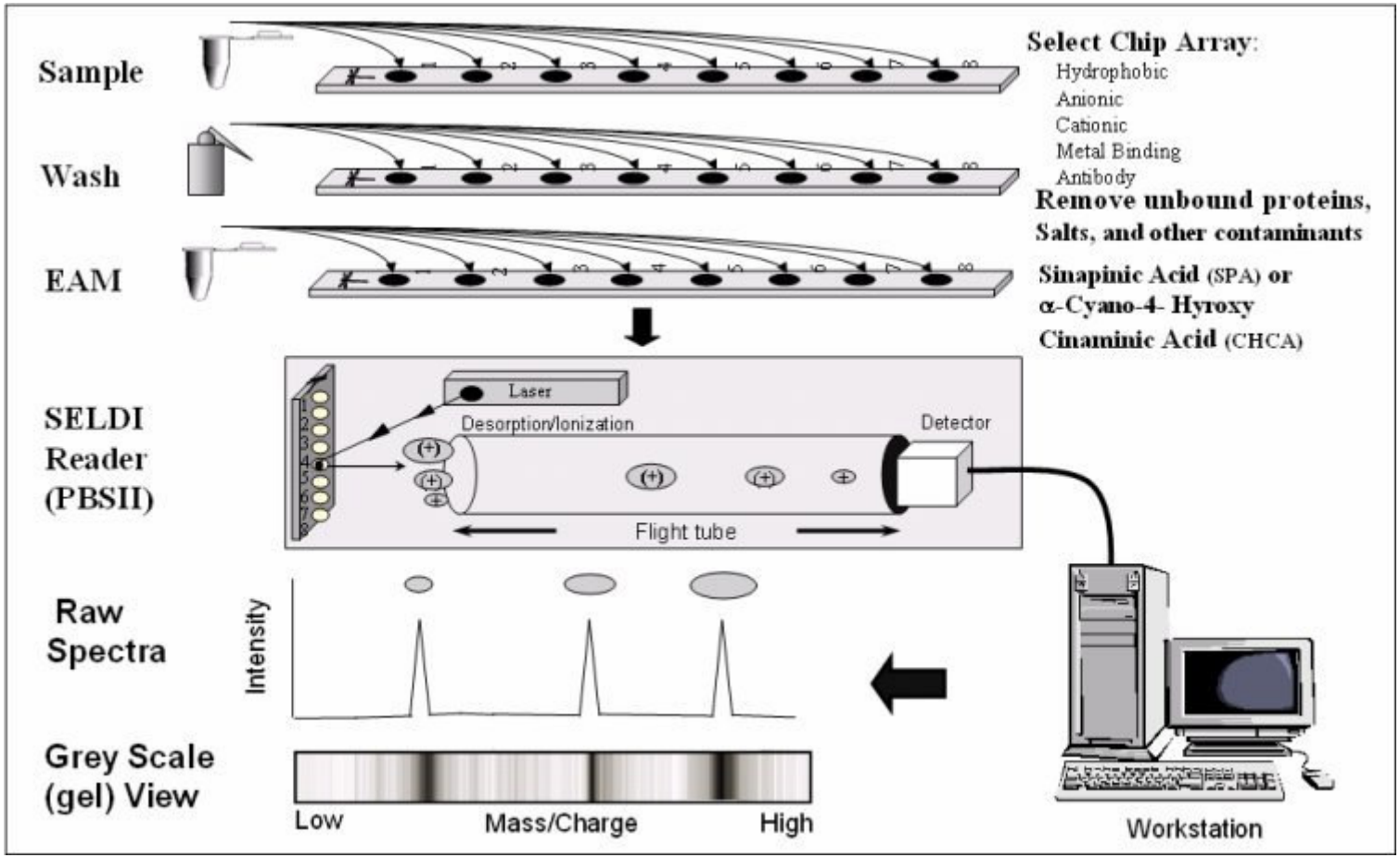
Writing a Manuscript Review

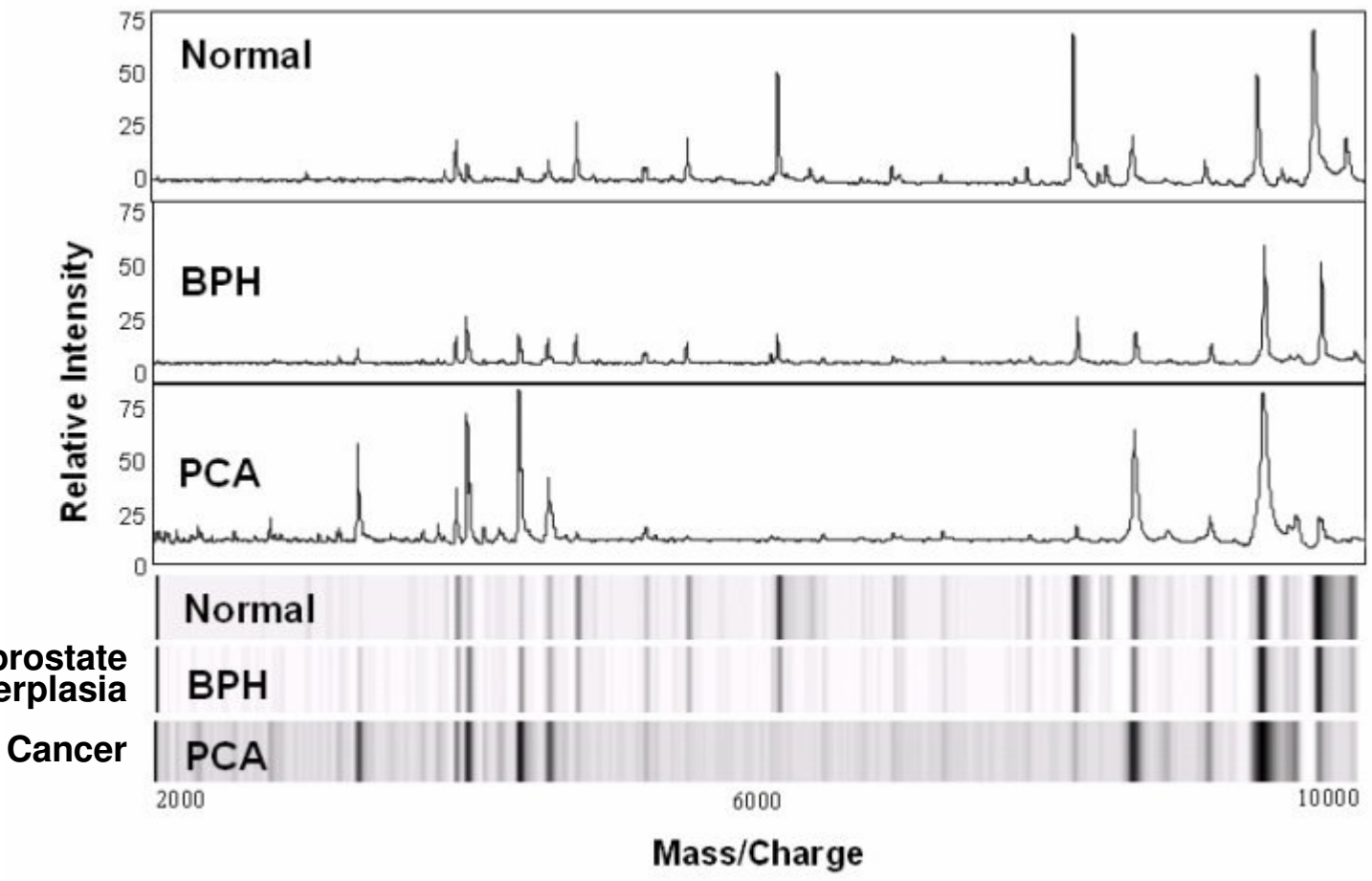
- Overview (~1-2 paragraphs) – summarizing the general point of the manuscript and your broad overall enthusiasm
- Major comments – bullet points
  - Items that must be fixed before publication
- Minor comments – bullet points
  - These should be stylistic recommendations
  - Typos
  - Etc…

SELDI ProteinChip™ Technology Process

- Sample goes *directly* onto the ProteinChip™ Array
- Proteins● are captured and *retained* on the chip (affinity capture ⬭)
- EAM△ is added to the chip
- Retentate map is "read" by Surface-Enhanced Laser Desorption/Ionization (SELDI)




Sample
Wash
EAM
Select Chip Array:
Hydrophobic
Anionic
Cationic
Metal Binding
Antibody
Remove unbound proteins, Salts, and other contaminants
Sinapinic Acid (SPA) or α-Cyano-4- Hyroxy Cinaminic Acid (CHCA)
SELDI Reader (PBSII)
Laser
Desorption/Ionization
Detector
Flight tube
Raw Spectra
Intensity
Grey Scale (gel) View
Low
Mass/Charge
High
Workstation


Normal
BPH
PCA
75
50
25
0
Relative Intensity
Normal
BPH
PCA
benign prostate hyperplasia
Prostate Cancer
2000
6000
10000
Mass/Charge