Strategies for Characterizing Proteins in Proteomics Research

Joseph A. Loo

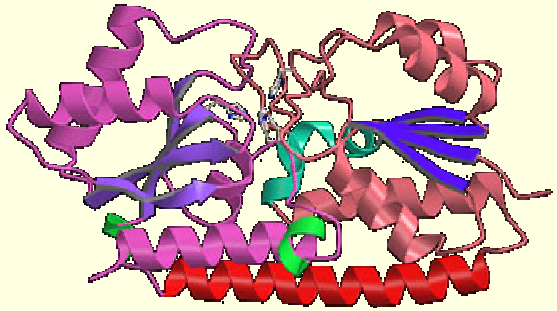
Department of Biological Chemistry
David Geffen School of Medicine

Department of Chemistry and Biochemistry

University of California
Los Angeles, CA USA

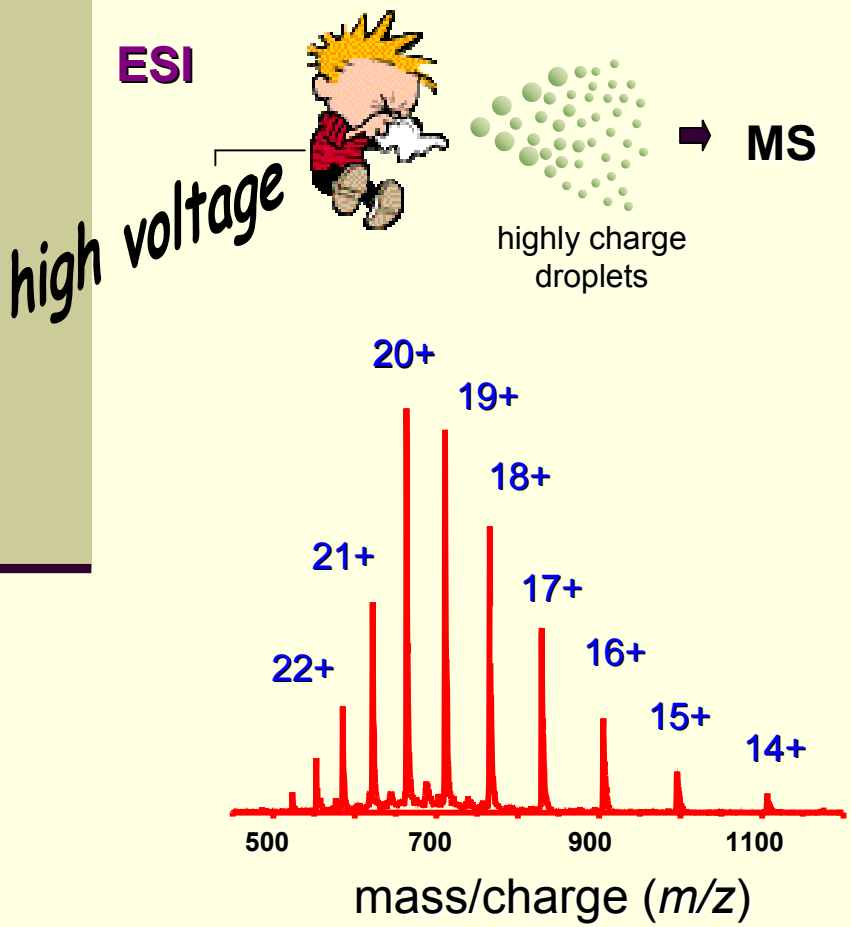
Approaches for Protein Identification

What is this protein?



- Molecular weight
- Isoelectric point (charge)
- Amino acid composition
- Other physical/chemical characteristics
- Partial or complete amino acid sequence
  - Edman (N-terminal sequence) - if N-term. not blocked
  - C-terminal sequence - not commonly performed
  - Mass spectrometry-measured information

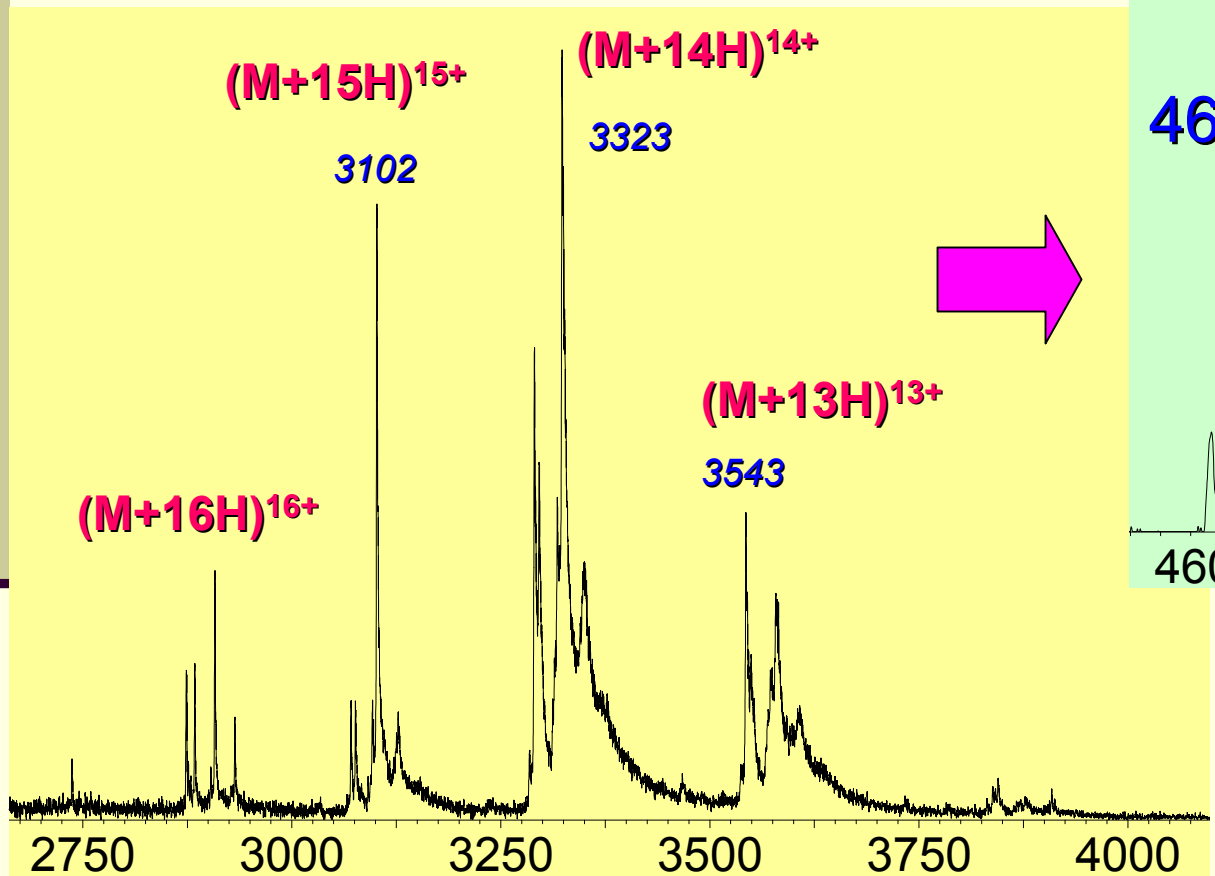
Electrospray Ionization (ESI)

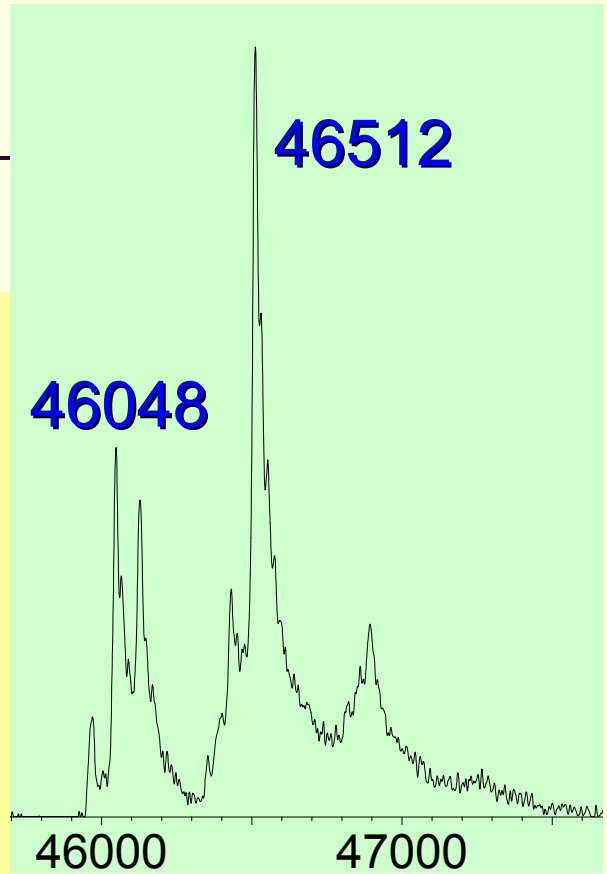


- Multiple charging
  - More charges for larger molecules
- MW range > 150 kDa
- Liquid introduction of analyte
  - Interface with liquid separation methods, e.g. liquid chromatography
  - Tandem mass spectrometry (MS/MS) for protein sequencing

ESI-MS of Large Proteins

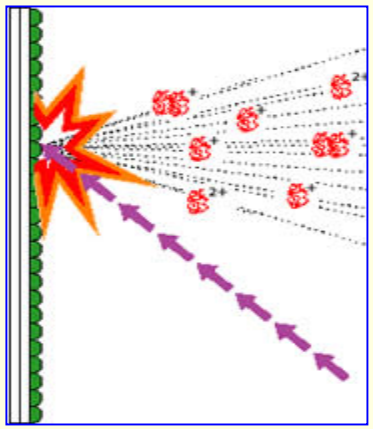
distribution of multiply charged molecules





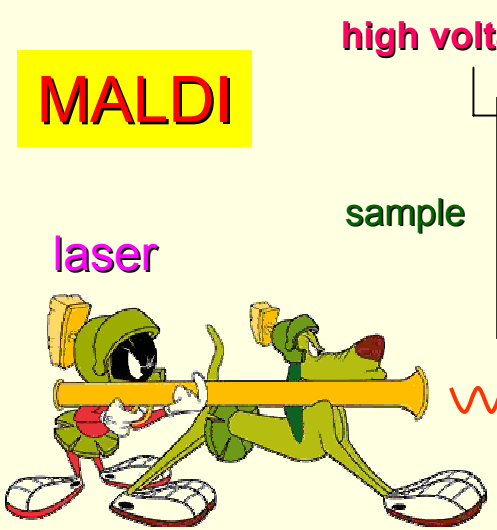
ESI-MS (Q-TOF)
pH 7.5

Matrix-assisted Laser Desorption/Ionization (MALDI)

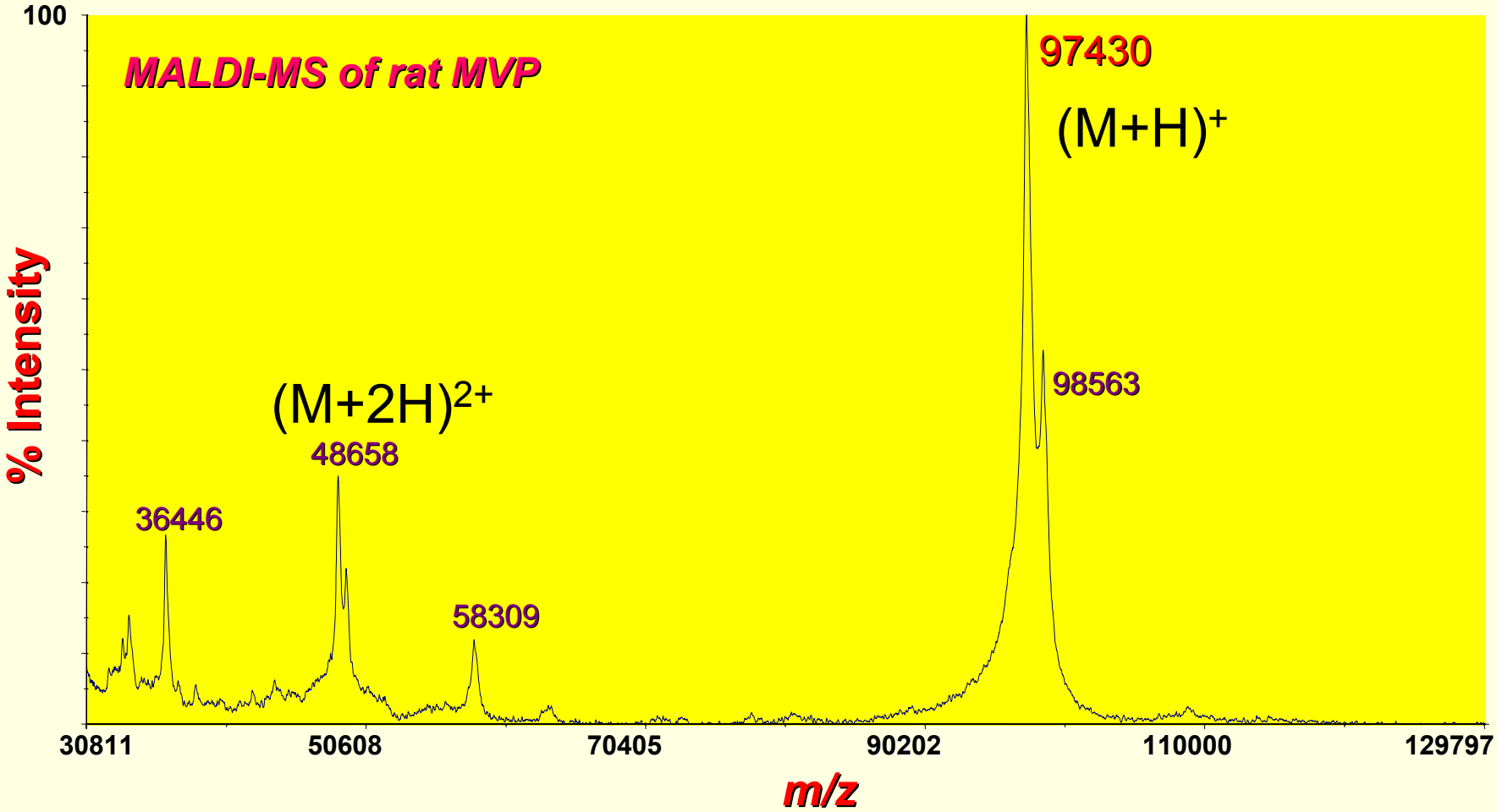


Time-of-Flight (TOF) Analyzer

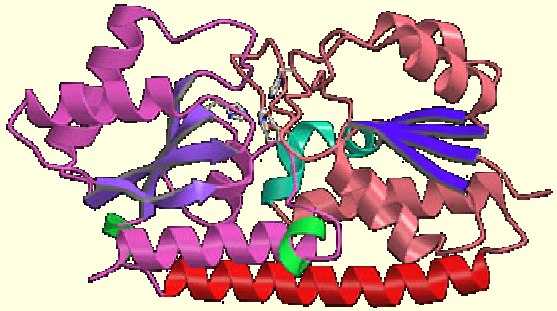
MALDI



MALDI Mass Spectrometry of Large Proteins



Approaches for Protein Sequencing and Identification



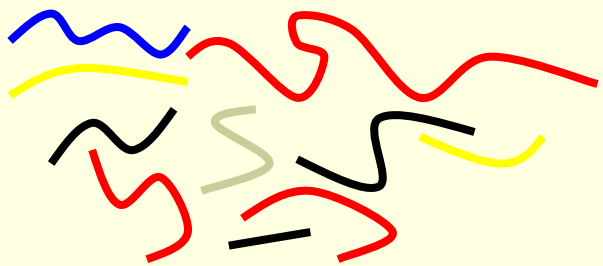
"Top Down"

MS/MS

MIRERICACVLALGMLTGFTHAFGSKDAAADGKPLVVTTIGMIADAVKNIAQGDVHLKGLMGP
GVDPHLYTATAGDVEWLGNADLILYNGLHLETKMGEVFSKLRGSRLVVAVSETIPVSQRLSLE
EAEFDPHVWFDVKLWSYSVKAVYESLCKLLPGKTREFTQRYQAYQQQLDKLDAYVRRKAQS
LPAERRVLVTAHDAFGYFSRAYGFEVKGLQGVSTASEASAHDMQELAAFIAQRKLPAIFIESSI
PHKNVEALRDAVQARGHVVQIGGELFSDAMGDAGTSEGTYVGMVTHNIDTIVAALAR

MS/MS

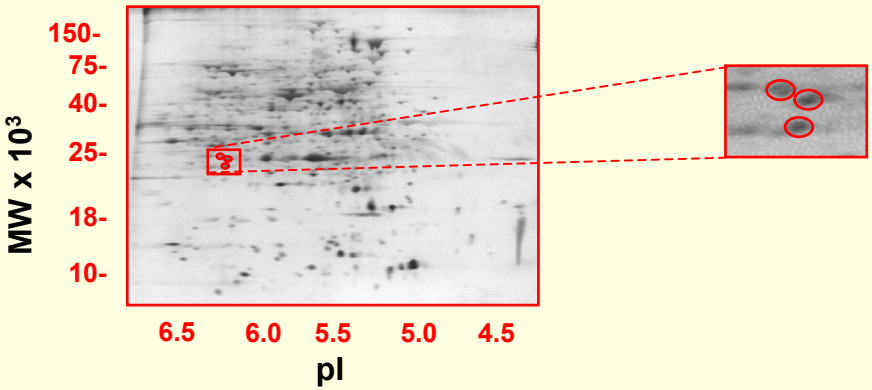
Enzymatic or chemical degradation



"Bottom Up"

Protein Identification by Mass Spectrometry

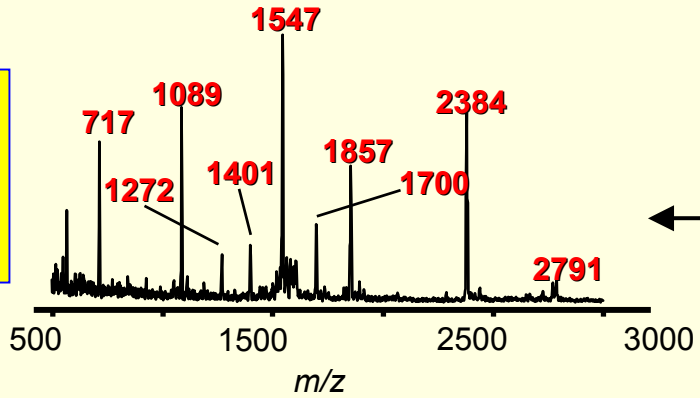
2-D Gel Electrophoresis



Excise separated protein "spots"

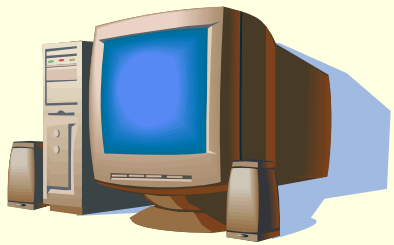
In-gel trypsin digest

Recover tryptic peptides



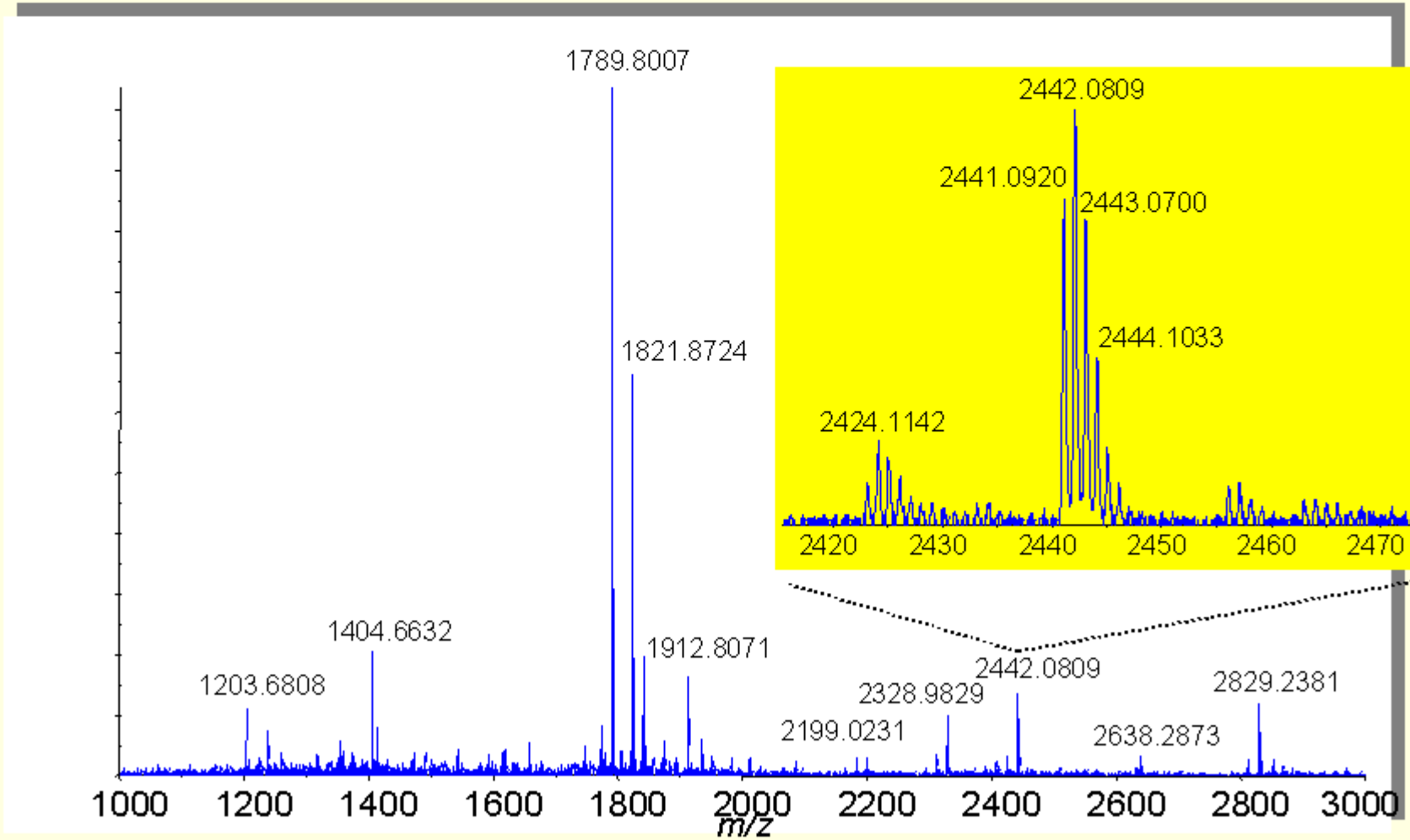
Peptide mass fingerprint by MALDI-TOF or LC-ESI-MS. Additional sequence information can be obtained by MS/MS.

Protein identification by searching proteomic or genomic databases



MALDI-QqTOF

Yeast Enolase (46 kDa) Tryptic Digest

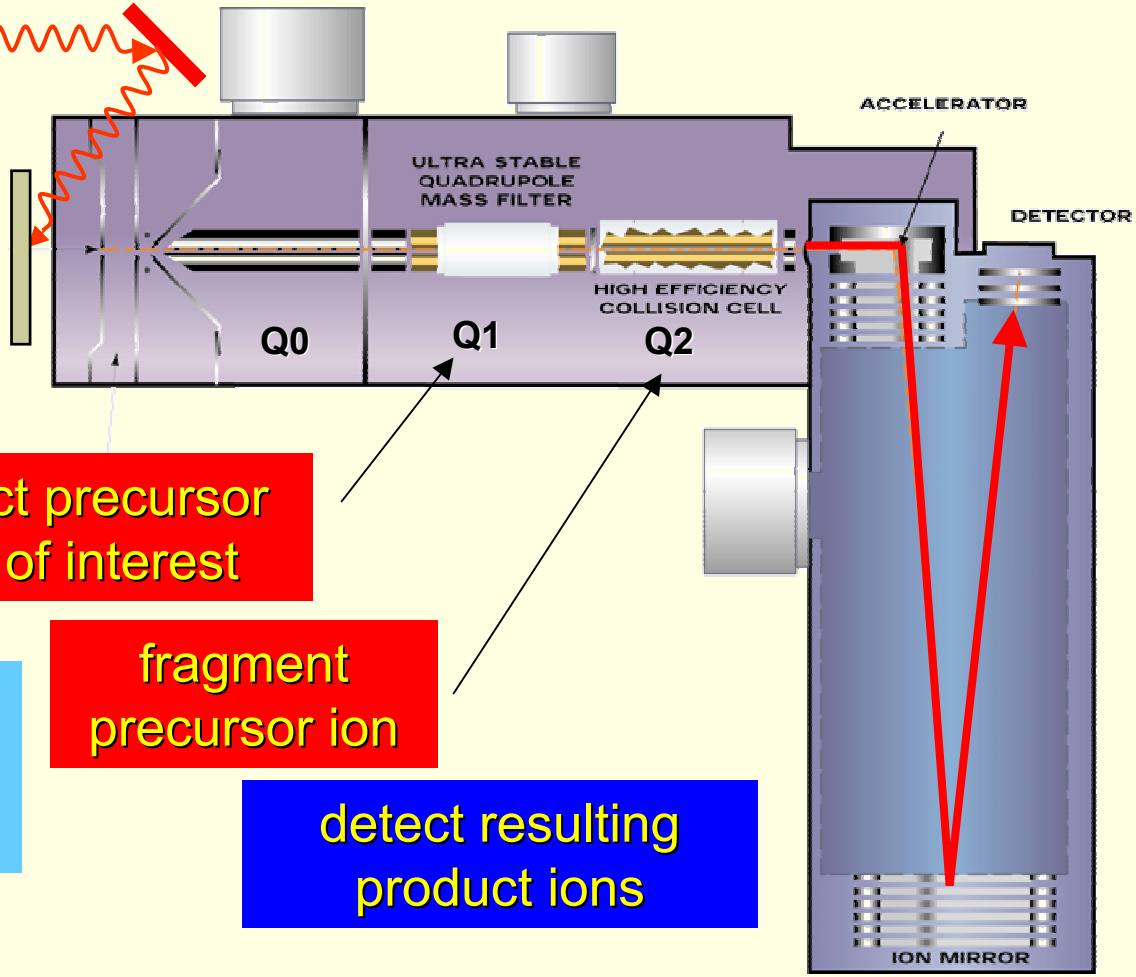


Quadrupole - Time-of-Flight (QTOF)

MALDI

ESI

high voltage



select precursor ion of interest

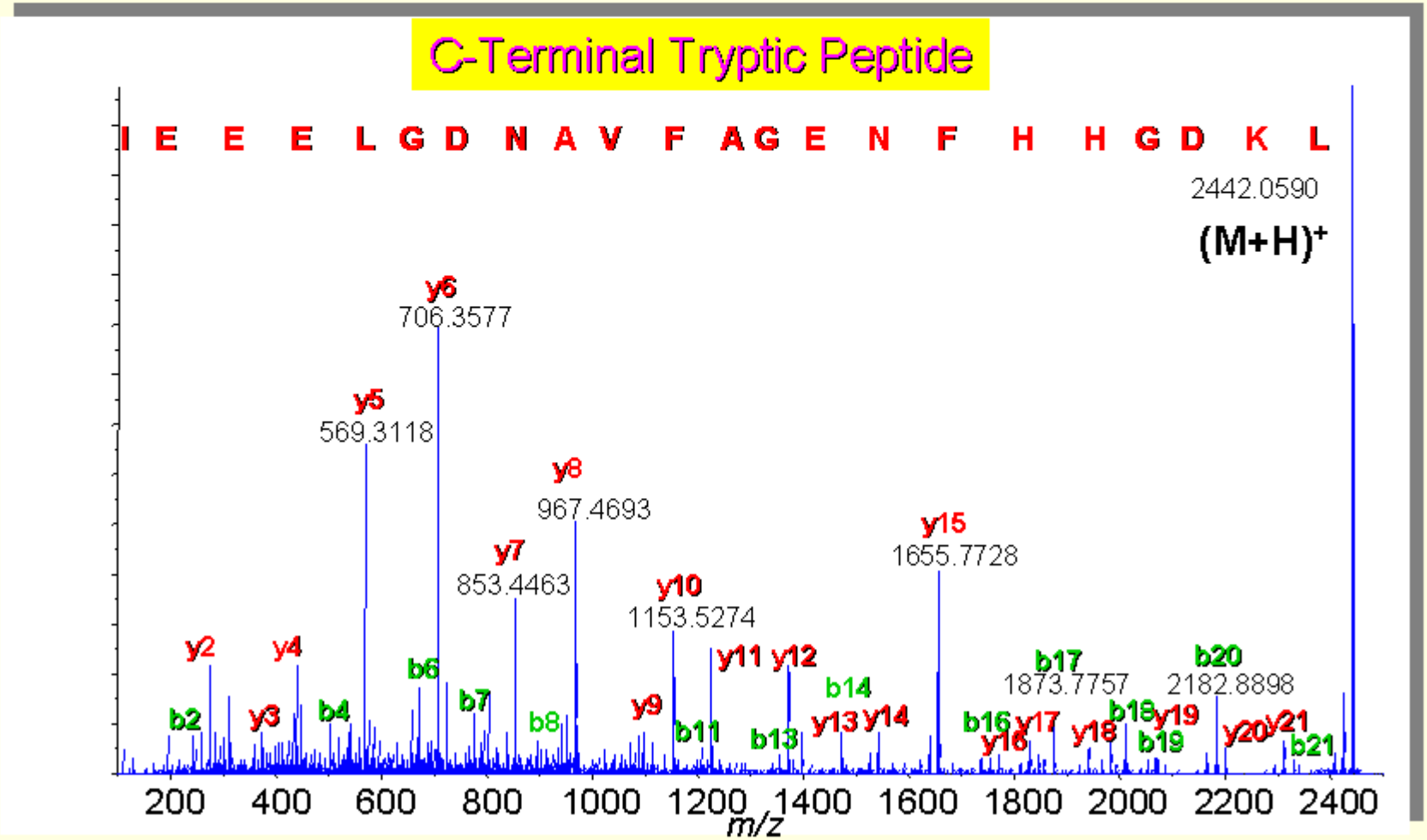
fragment precursor ion

detect resulting product ions

tandem mass spectrometry "MS/MS"

MALDI-QqTOF-MS/MS

Yeast Enolase (46 kDa) Tryptic Digest



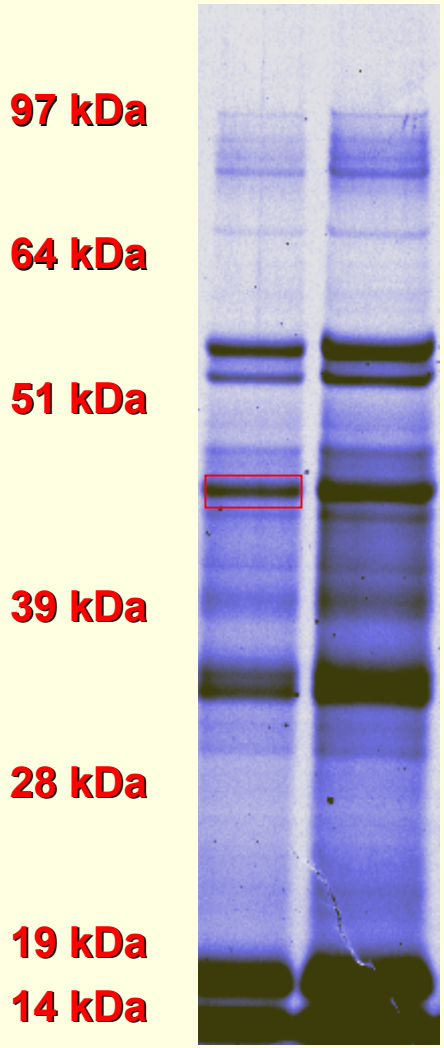
De Novo Protein Sequencing of Heat Resistant Proteins Found in Sacred Lotus

- Lotus, *Nelumbo nucifera*, has the distinction of having produced the longest living seed, 1300 yr
- Seed longevity is attributable to its
  - Fruit coat's impermeability to O_2 and water
  - Rapid adaptability to stress
- Genome is not known
  - de novo protein sequencing

Jane Shen-Miller, Kerry Wooding, Yongming Xie

SDS-PAGE

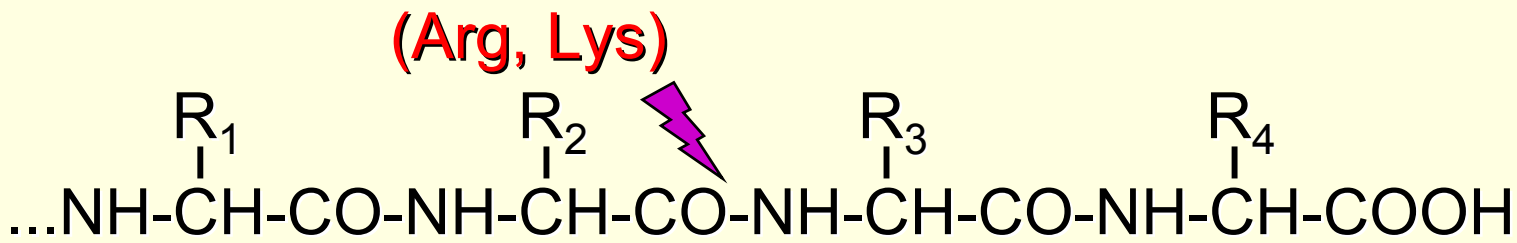
Lotus protein (>400 yr) boiled embryo extracts

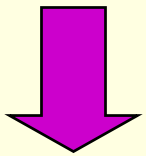


- Heat hardy proteins are soluble in >100° C
  - enable draught protection
  - Lotus Fe-SOD (oxidative repair) retains 65% activity after 1 hr boiling
  - mass spectrometry and de novo sequencing to identify heat stable proteins

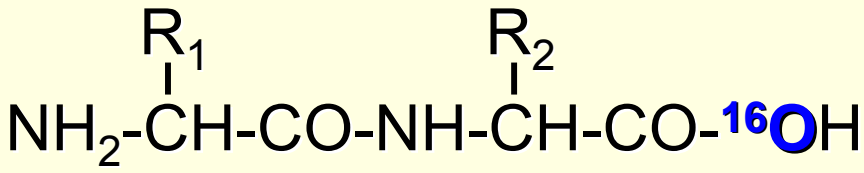
Isotope Labeling as an Aid to *De Novo* Sequencing

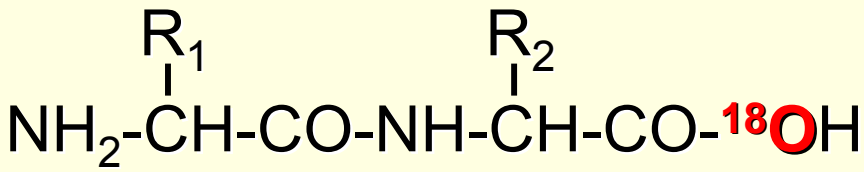
Trypsin Digestion in $H_2{}^{18}O$





Trypsin / $H_2{}^{16}O$ / $H_2{}^{18}O$





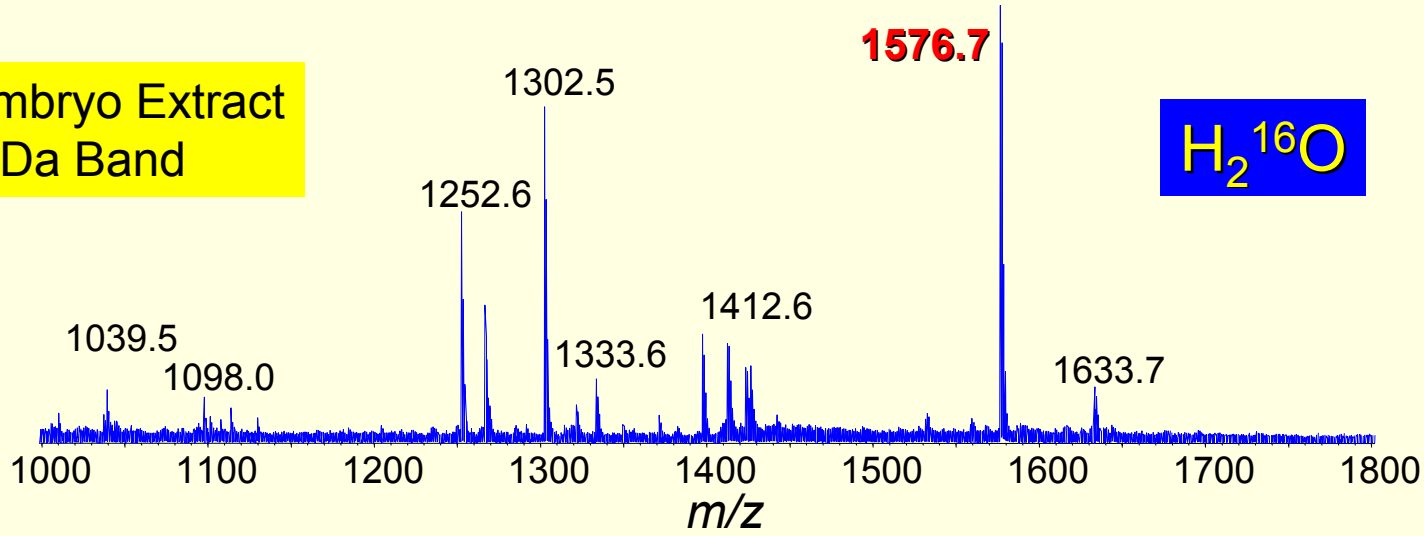
^{16}O or ^{18}O

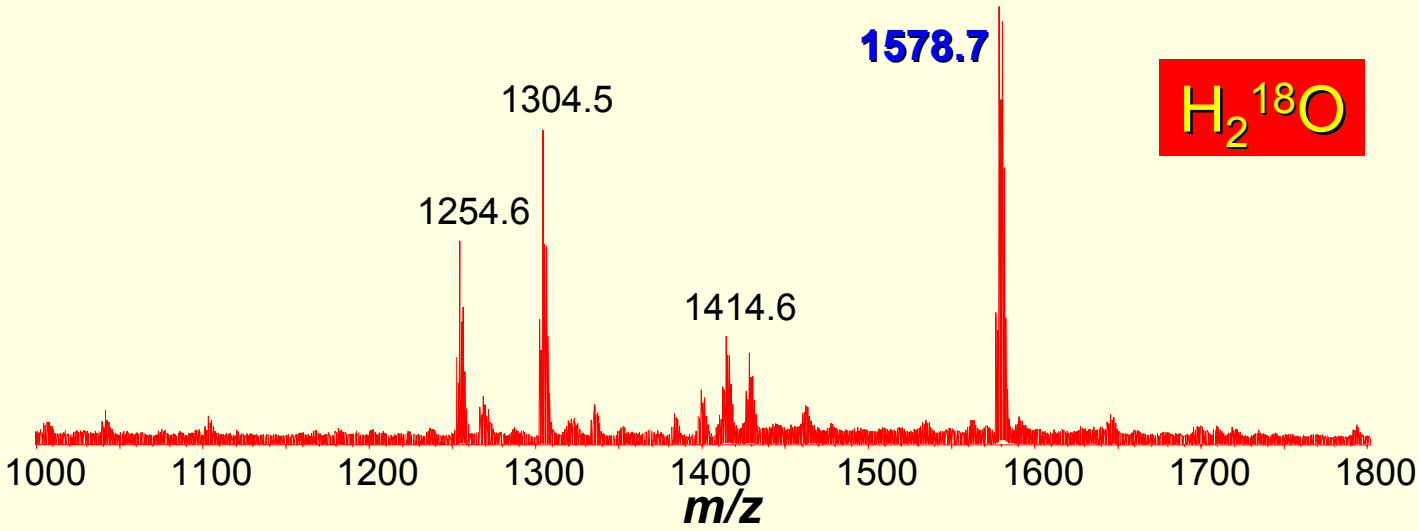
- Aids interpretation of MS/MS spectra
- Distinguish b-ions (from N-terminus) and y-ions (from C-terminus): **2 Da shift**

MALDI-QqTOF Tryptic Peptide Map

trypsin digestion performed in $H_2{}^{16}O$ or $H_2{}^{18}O$

Boiled Embryo Extract
45 kDa Band

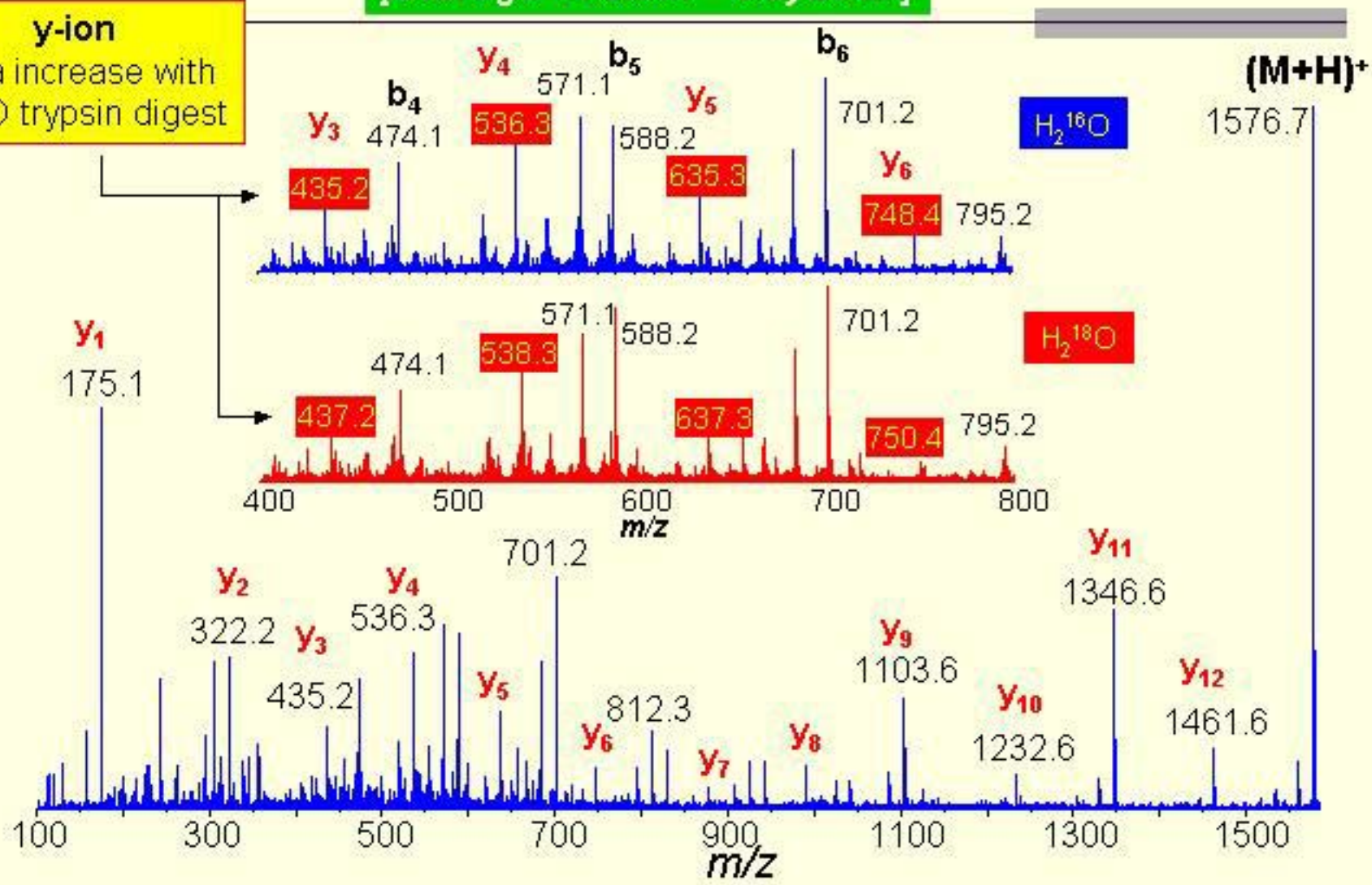




MALDI-QqTOF MS/MS

DDNEN(I/L)Q(I/L)VT(I/L)FR
[Storage Protein - Soybean]

y-ion
2 Da increase with
$H_2{}^{18}O$ trypsin digest

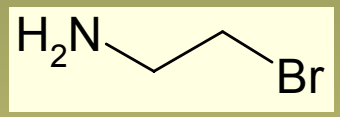


Enzyme and Chemical Specificity for Proteolysis

Arg/Lys - XX	Trypsin
Lys - XX	Lys-C
Arg - XX	Arg-C
Asp/Glu - XX	V-8 protease
XX - Asp	Asp-N
Met - XX	CNBr

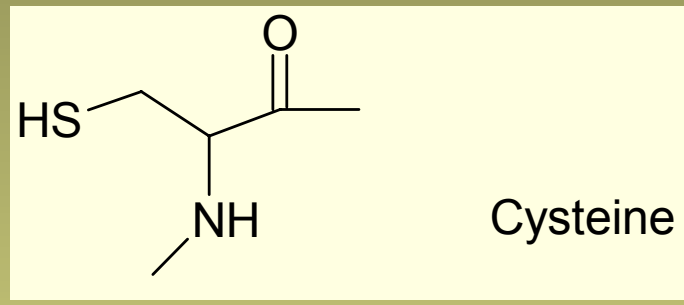
New Chemistries for Proteomics:

Chemical Modification of Cysteine (Cys-specific cleavage)

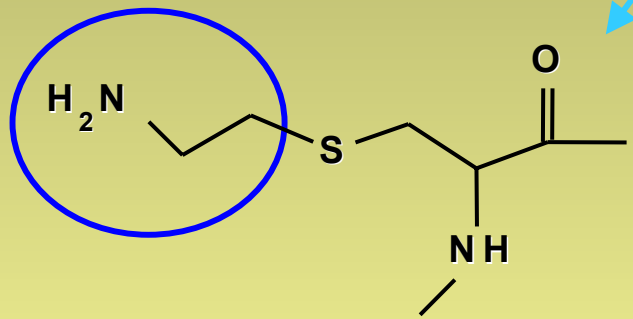


2-bromoethyl amine

OH^-



Lindley, H. *Nature*, 1956, *178*, 647.



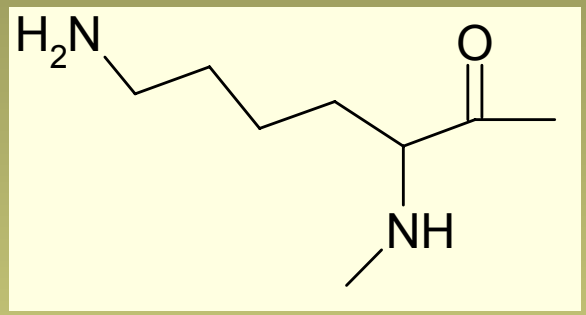
S-aminoethyl cysteine

Digest with trypsin to cleave after Lys, Arg, Cys

Thevis and Loo
J. Proteome Res., 2003

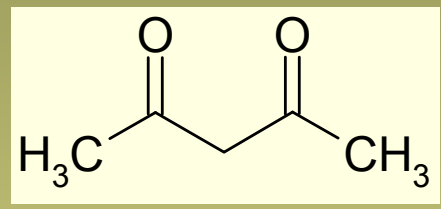
Chemical Modification of Lysine

Block Lys-cleavage



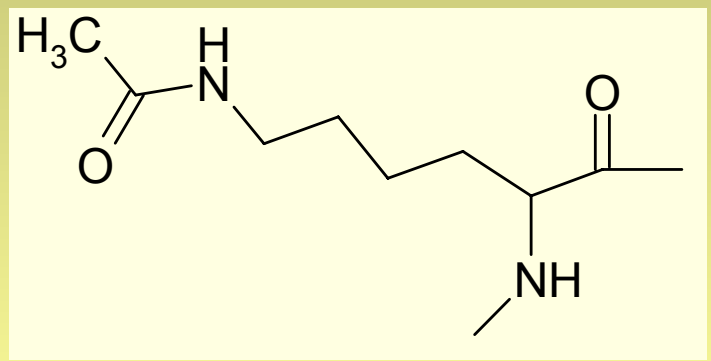
Lysine

+



Acetic anhydride

MeOH



Lysine, acetylated (+ 42 Da)

Digest with trypsin to cleave after Arg, Cys

Digest with Lys-C to cleave after Cys

Chemical Modification of Cysteine

Cys-specific cleavage

Ionization enhancement of Cys-peptides

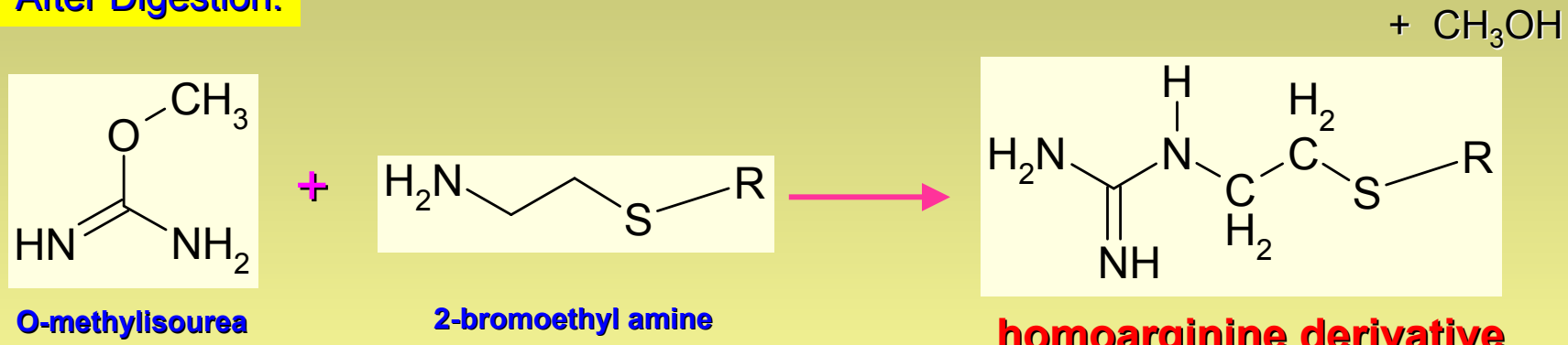
Prior to Digestion:

2-bromoethyl amine

S-aminoethyl cysteine

Lindley, H. *Nature*, **1956**, *178*, 647.

After Digestion:



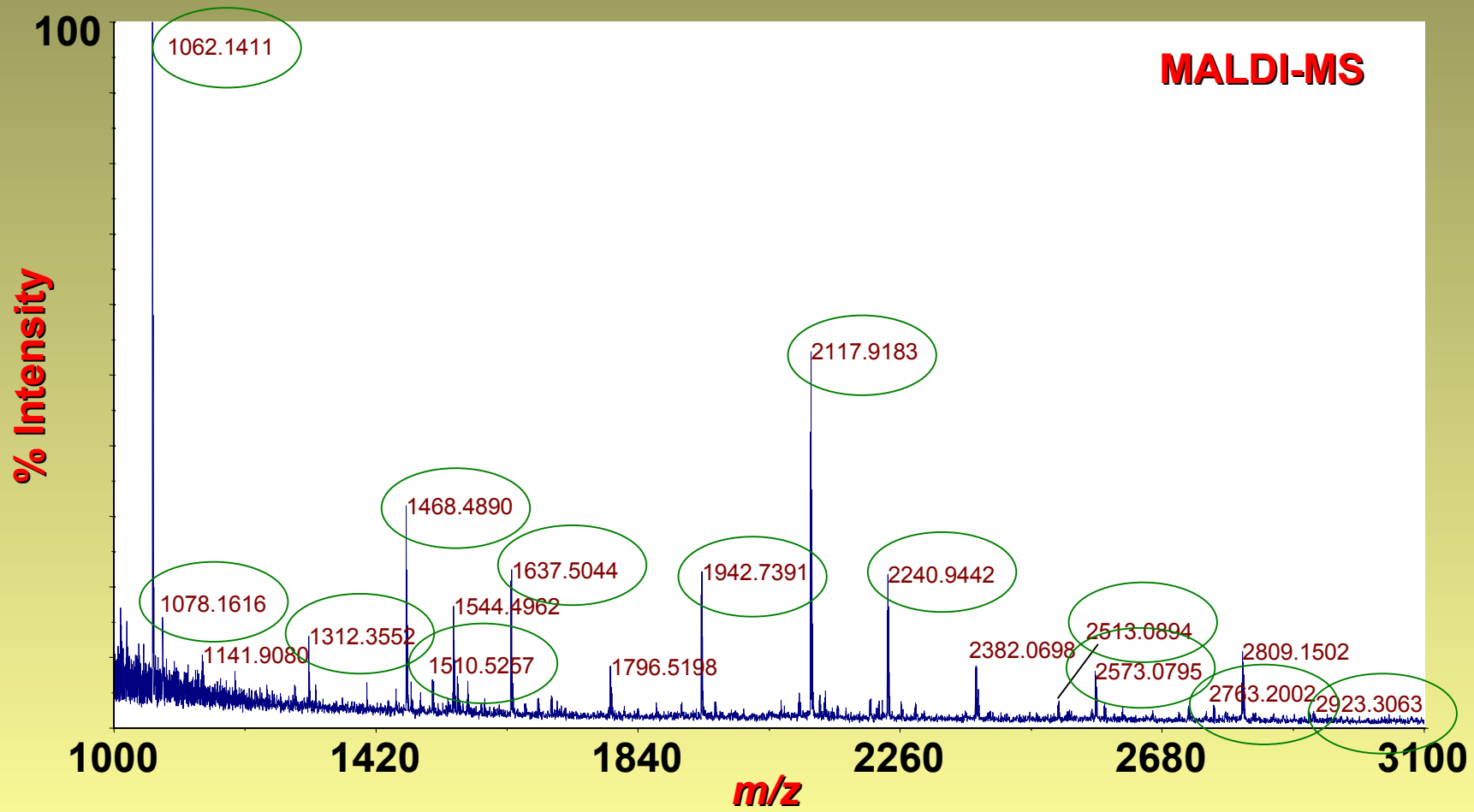
O-methylisourea

2-bromoethyl amine

homoarginine derivative (improve MALDI response)

Hale, J.E. et al. *Anal. Biochem.* **2000**, *287*, 110.

Human Serum Albumin
In-gel derivatization and trypsin digestion



Human Serum Albumin
In-gel derivatization and trypsin digestion (cleavage after Cys, Arg)

DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA
KTCVADESAE NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE
CFLQHKDDNP NLPRLVRPEV DVMCTAFHDN EETFLKKYLY EIARRHPYFY
APELLFFAKR YKAAFTECCQ AADKAACLLP KLDELRDEGK ASSAKQRLKC
ASLQKFGERA FKAWAVARLS QRFPKAEFAE VSKLVTDLTK VHTECCHGDL
LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA
DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVVLLLRLA
KTYETTLEKC CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE
YKFQNALLVR YTKKVPQVST PTLVEVSRNL GKVGSKCCKH PEAKRMPCAE
DYLSVVLNQL CVLHEKTPVS DRVTKCCTES LVNRRPCFSA LEVDETYVPK
EFNAETFTFH ADICTLSEKE RQIKKQTALV ELVKHKPKAT KEQLKAVMDD
FAAFVEKCCK ADDKETCFAE EGKKLVAASQ AALGL

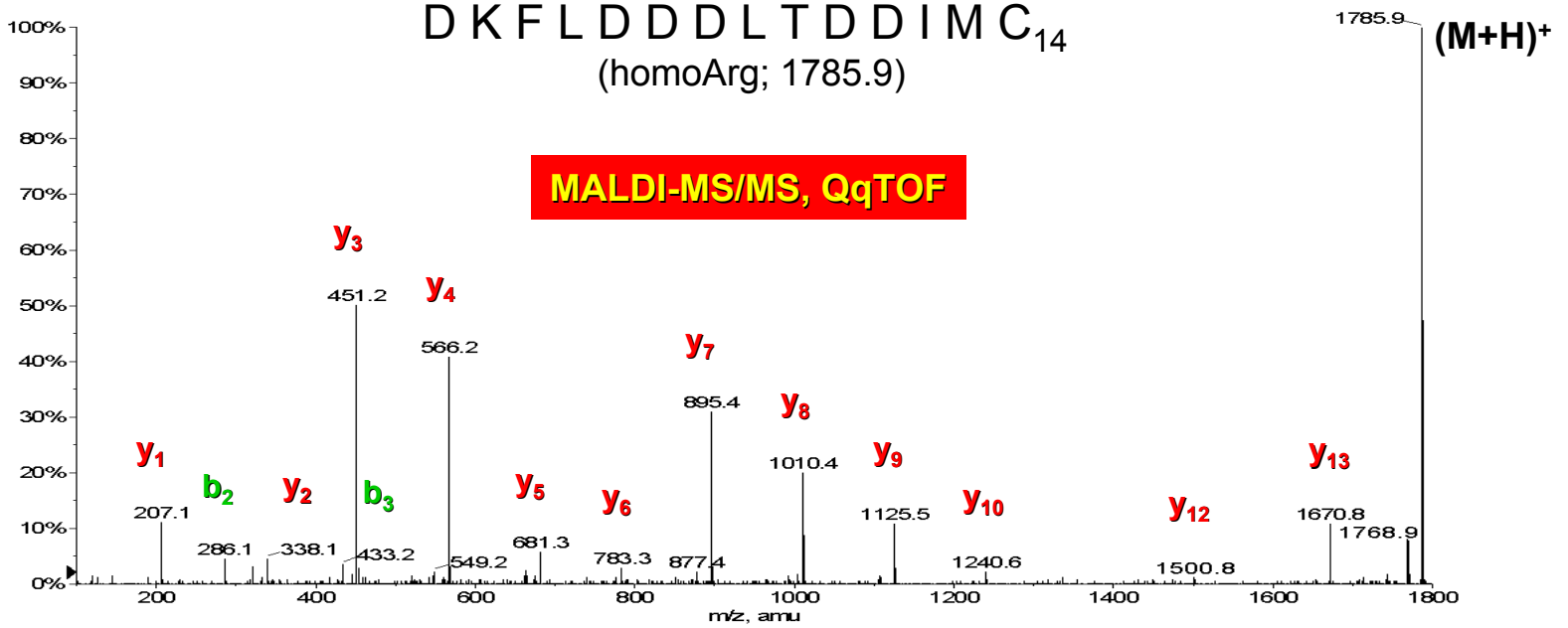
Without homoarginine derivatization
In addition, with homoarginine derivatization

α-Lactalbumin

In-gel derivatization and trypsin digestion
Homoarginine derivatization

EQLTKCEVFR ELKDLKGYGG VSLPEWVCTT FHTSGYDTQA IVQNNDSTEY
GLFQINNKIW CKDDQNPHSS NICNISCDKF LDDDLTDDIM CVKKILDKVG
INYWLAHKAL CSEKLDQWLC EKL

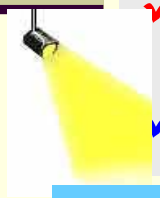
86% sequence coverage



Why Bother with Intact Protein Masses?

The protein's molecular mass defines the native covalent state of a gene's product, including:

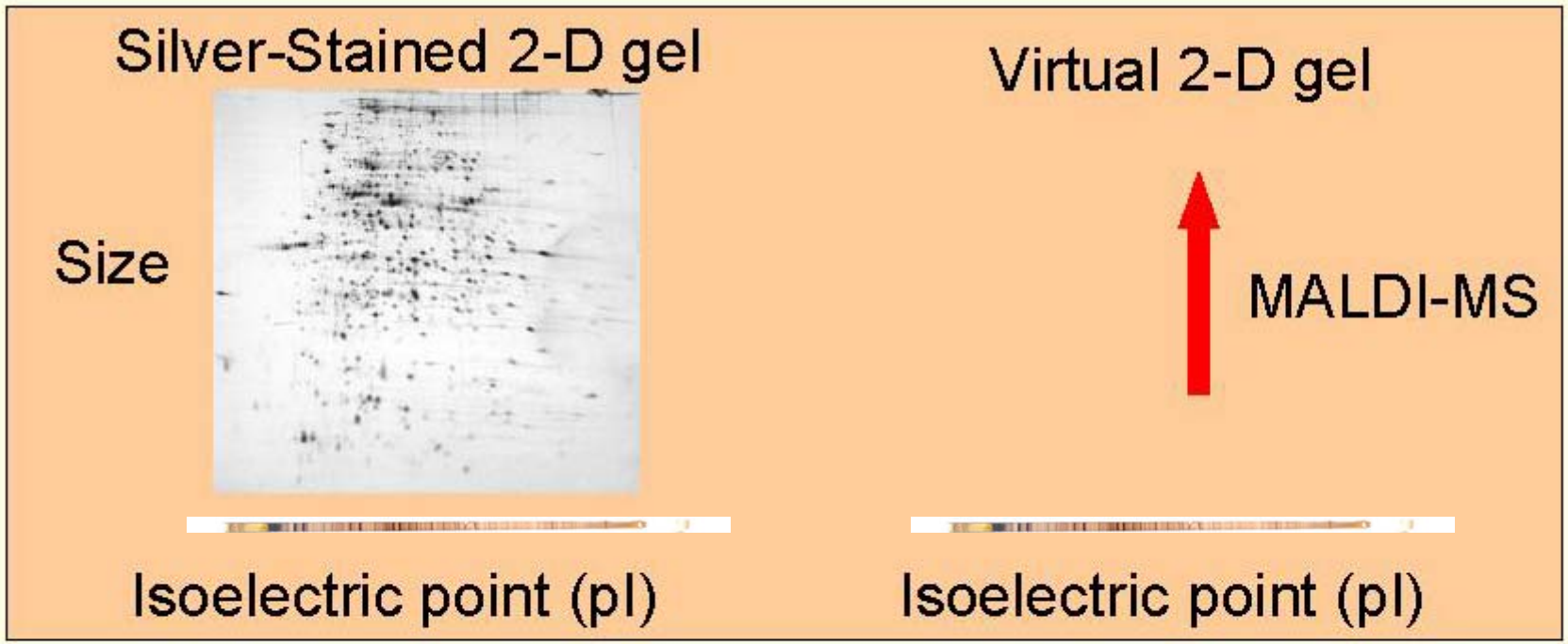
- ✓Post-translational modifications: Glycosylation, Phosphorylation, Oxidation, Deamidation, Lipoylation, Acetylation, Formylation...
- ✓Post-transcriptional modifications
- ✓Alternative splicing, Introns/Exons, Frame-shifts, Sequencing Errors
- ✓Simple confirmation of suspected id's
- ✓Spotlight the presence of post-translational modifications



The fragmentation pattern from proteins can generate sufficient information for identification from sequence databases, particularly when combined with accurate mass measurements of both the intact molecule and its product ions.

Combining Gel Electrophoreisis with Mass Spectrometry

- Building better tools
  - Profiling proteins faster in an automated fashion



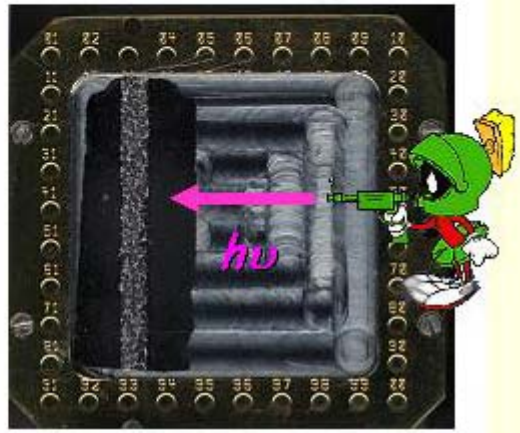
PAGE-MALDI-MS: *The Virtual 2D Gel Method*

Potential for automated, high-speed proteomic analysis

Separate protein mixture by IEF-PAGE (immobilized pH gradient gel, IPG)

Wash gel and soak gel in MALDI matrix solution

Mount gel onto MALDI target

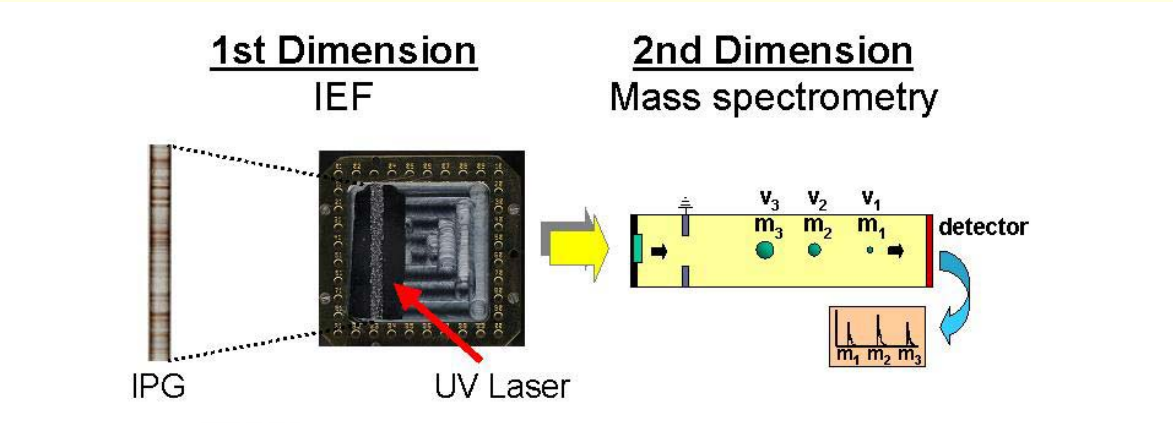


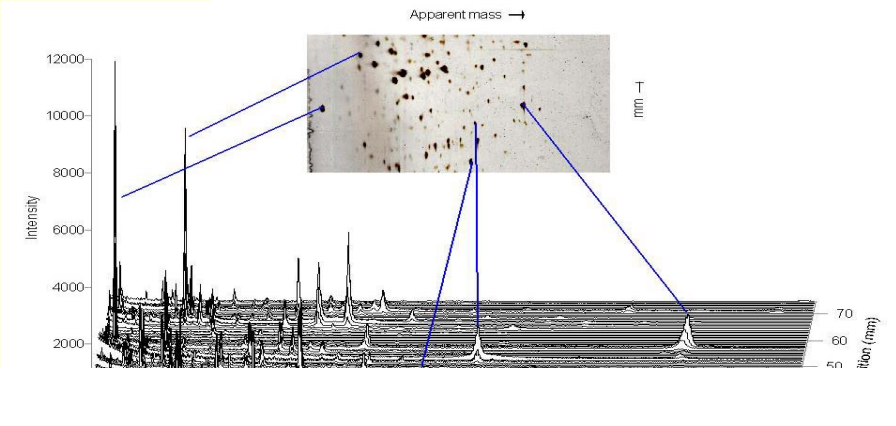
MALDI-TOF Mass Spectrometry

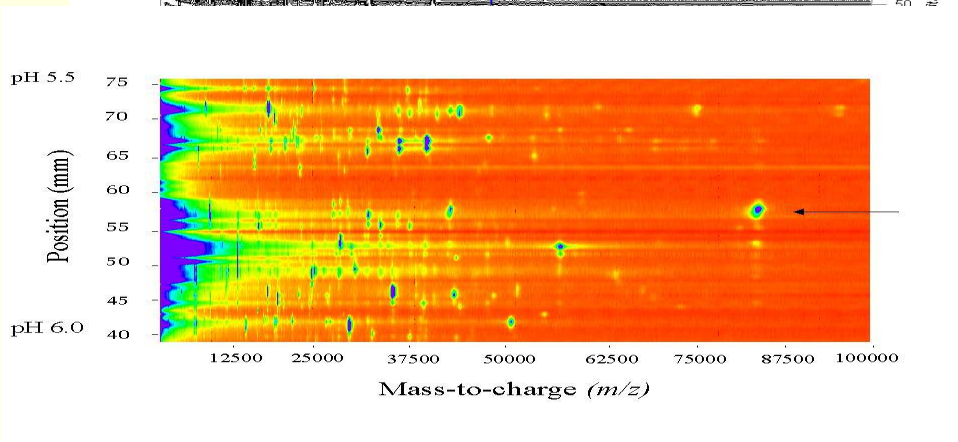
Rachel Loo

Virtual 2-D Gel Electrophoresis

(*Analytical Chemistry* 73, 4063 (*2001*))

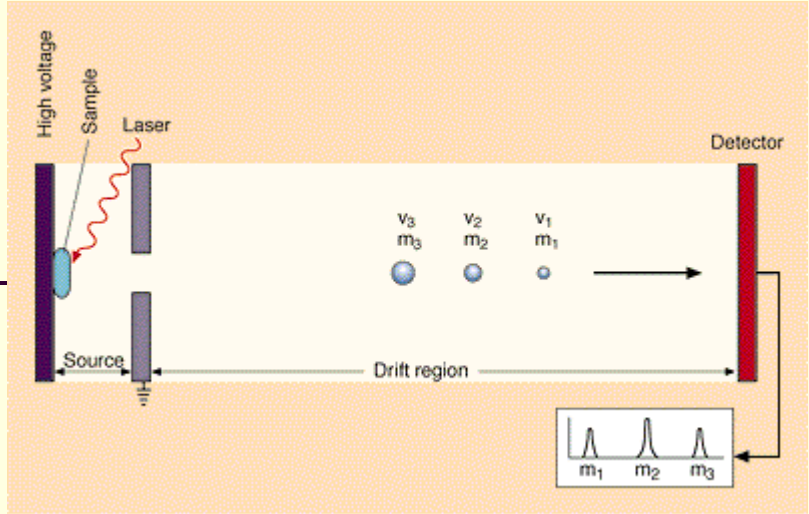




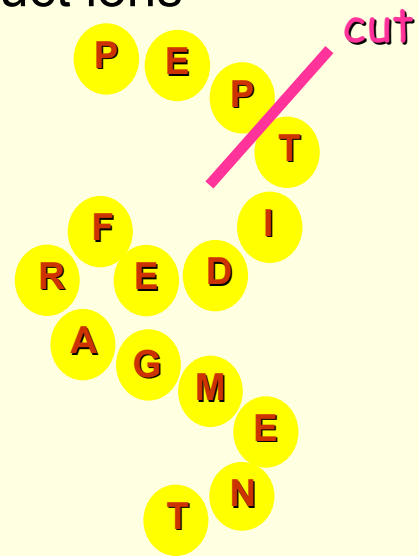


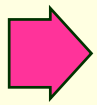
- Identify proteins based on MW and pI
- Create a database of intact protein masses and provide a direct link to all data produced now and in the future via classical 2D gels
- Identify proteome-wide post-translational modifications

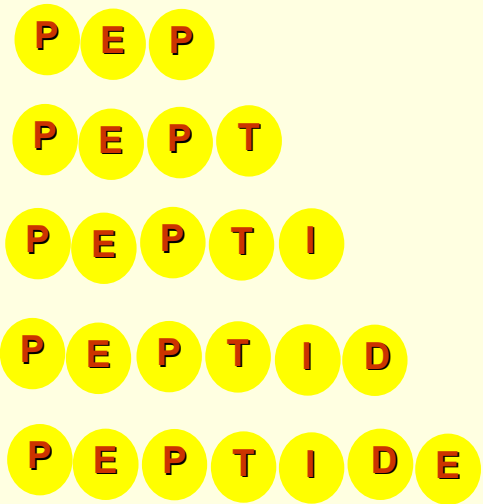
In-Source Decay (ISD) for Protein Sequencing



- Peptides and large proteins can be fragmented by ISD
- Fragmentation occurs in the MALDI ion source
  - not generally well controlled
- Reflectron TOF not necessary (linear TOF sufficient to measure product ions



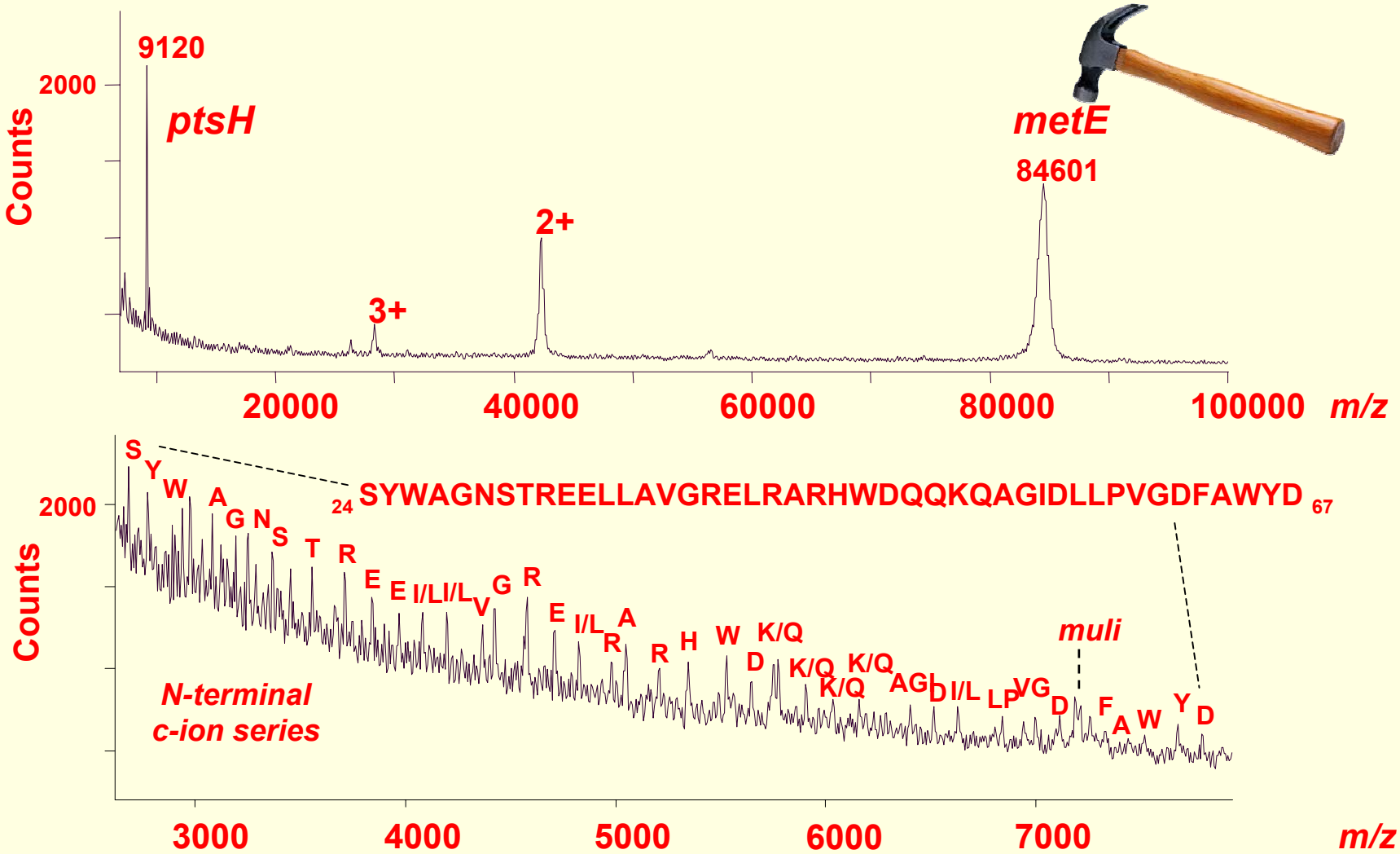




- Complete sequence information not present, but extensive stretches of sequence from the N- and/or C-termini observed

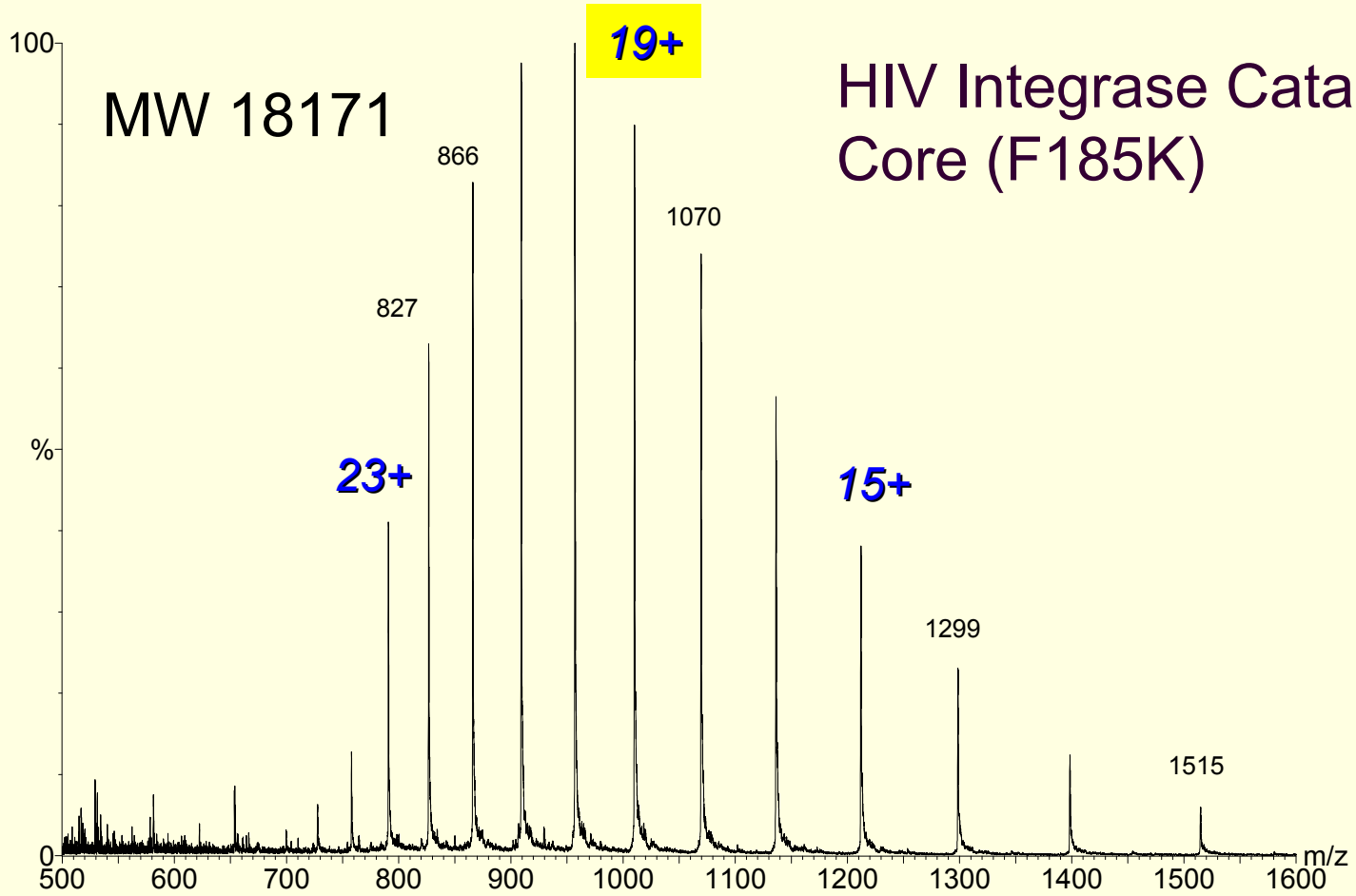
...TIDE...

Protein ISD from Isoelectric Focusing Gel

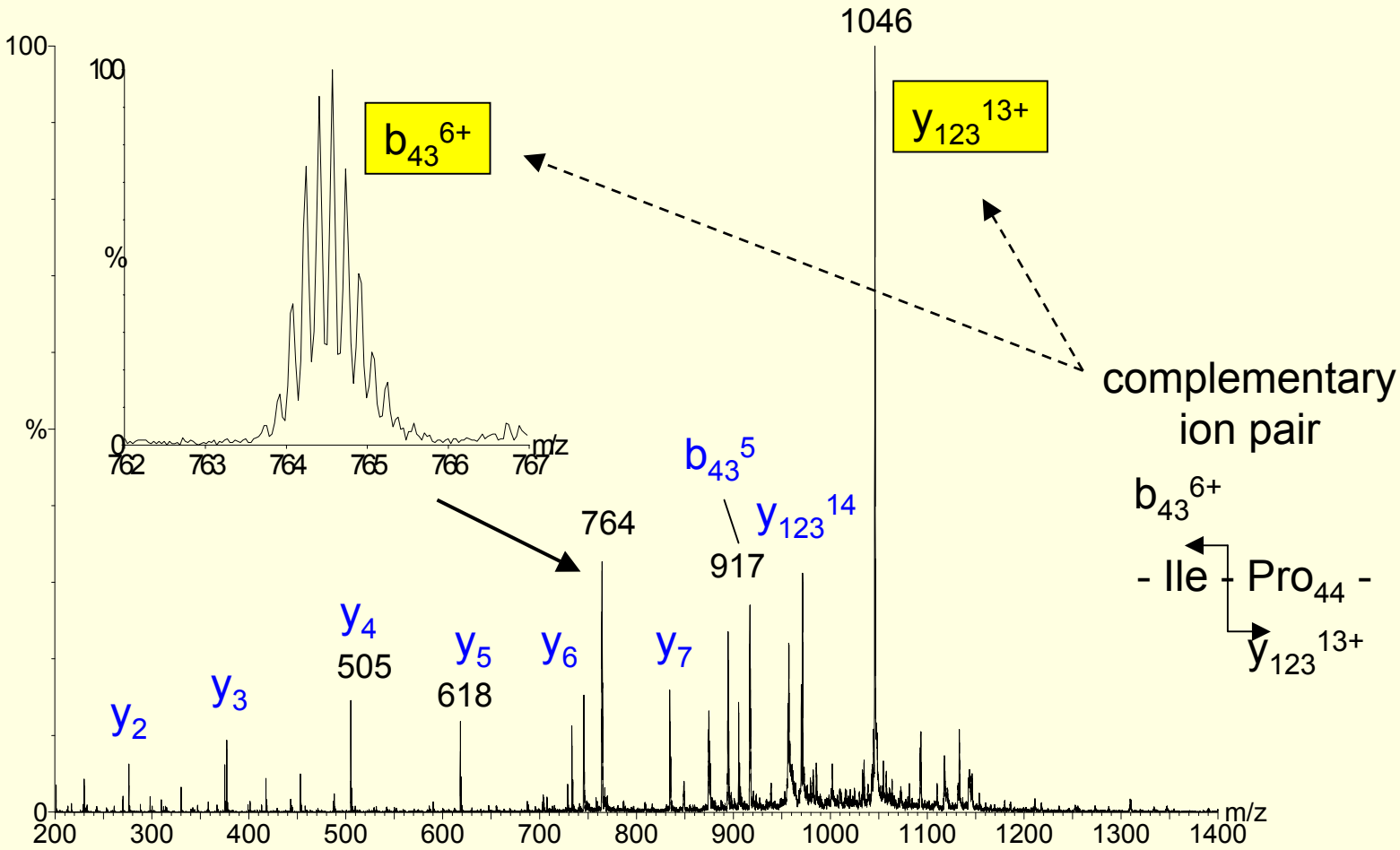


Ogorzalek Loo, RR, Cavalcoli, JD, VanBogelen, RA, Mitchell, C, Loo, JA, Moldover, B., and Andrews, PC, *Anal. Chem.* 73, 4063 (*2001*).

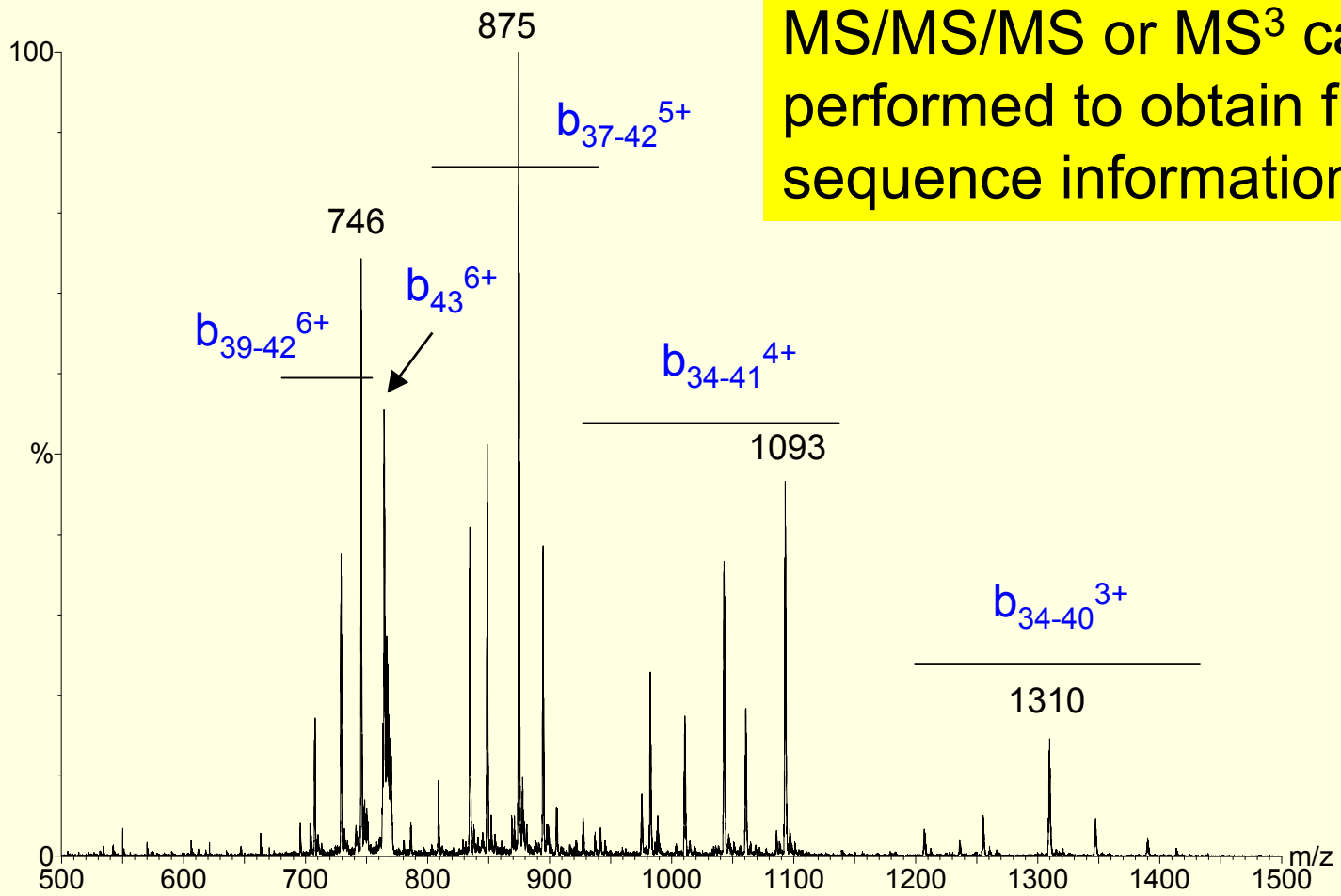
Top-down sequencing of proteins by ESI-MS/MS



Integrase: MS/MS $(M+19H)^{19+}$



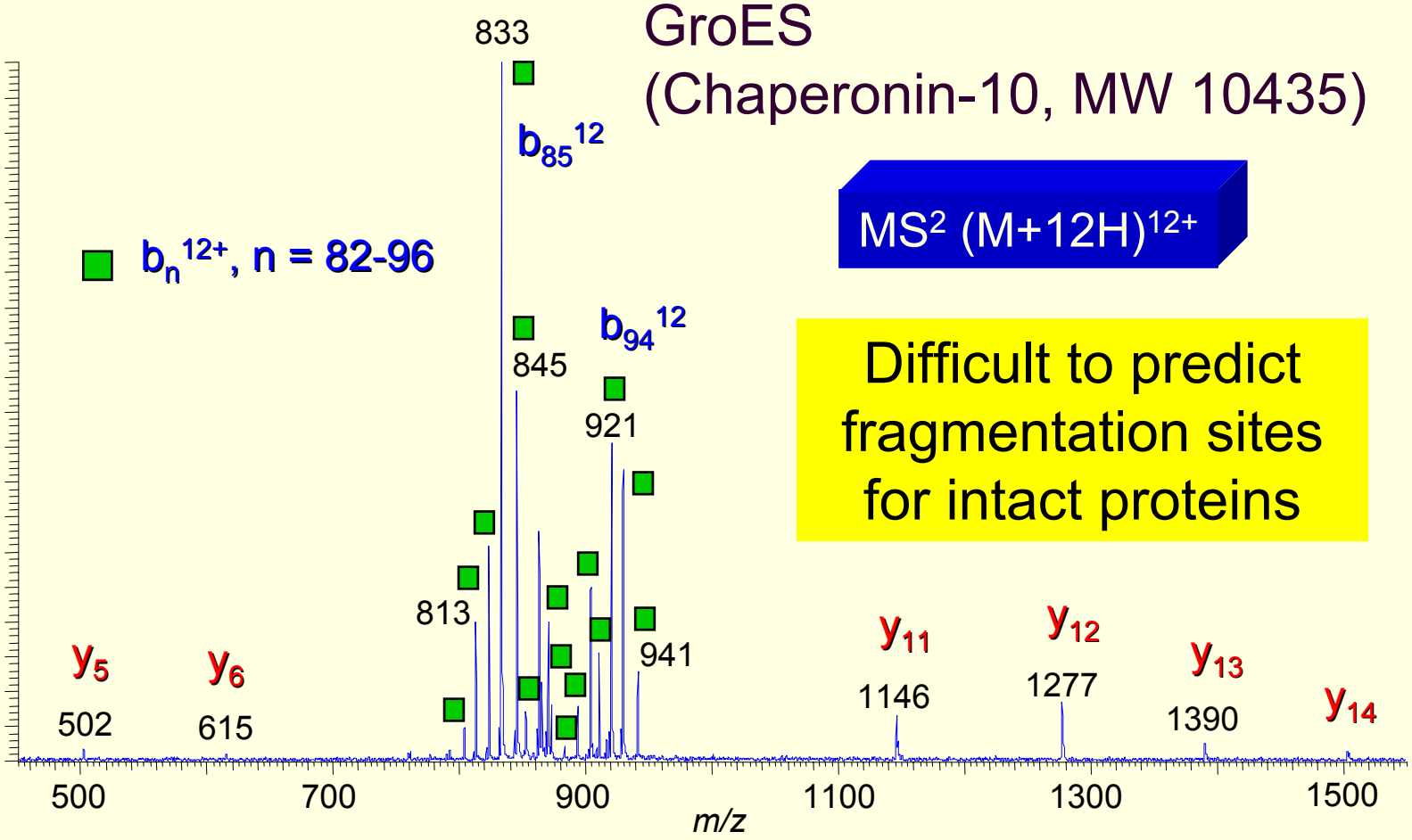
MS/MS b_{43}^{6+} (*m/z* 764)



MS/MS/MS or MS^3 can be performed to obtain further sequence information

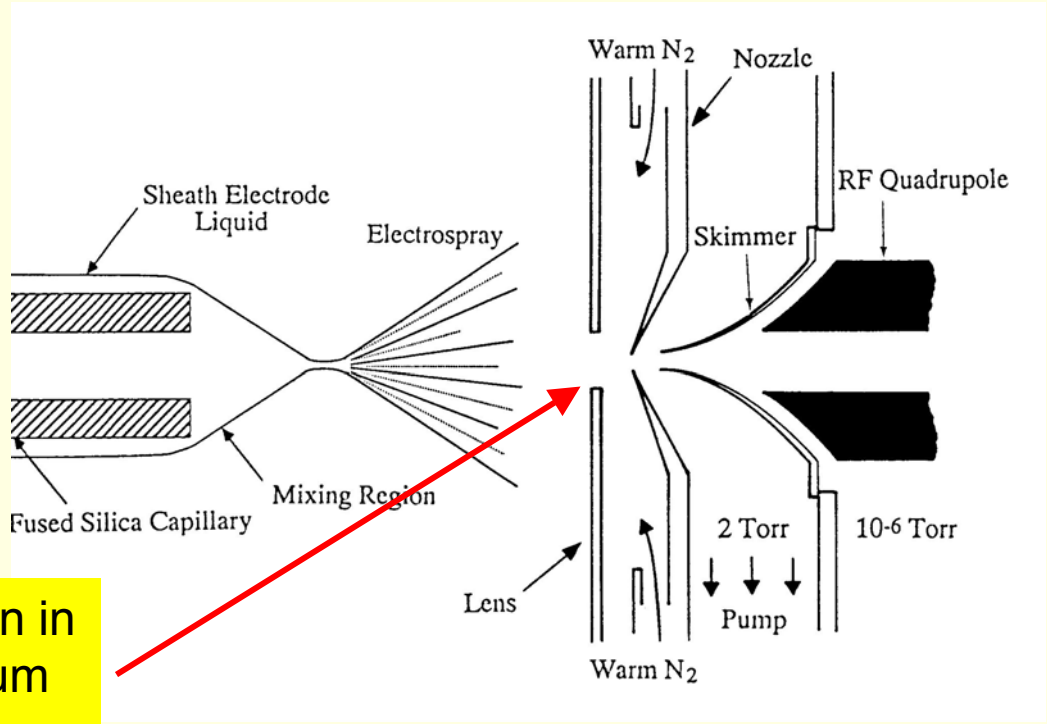
ESI-MS/MS of GroES (97 amino acids)

MS/MS yields direct sequence information for the C-terminal 15 amino acid residues.



Peptide fragmentation in “nozzle-skimmer” region of ESI interface

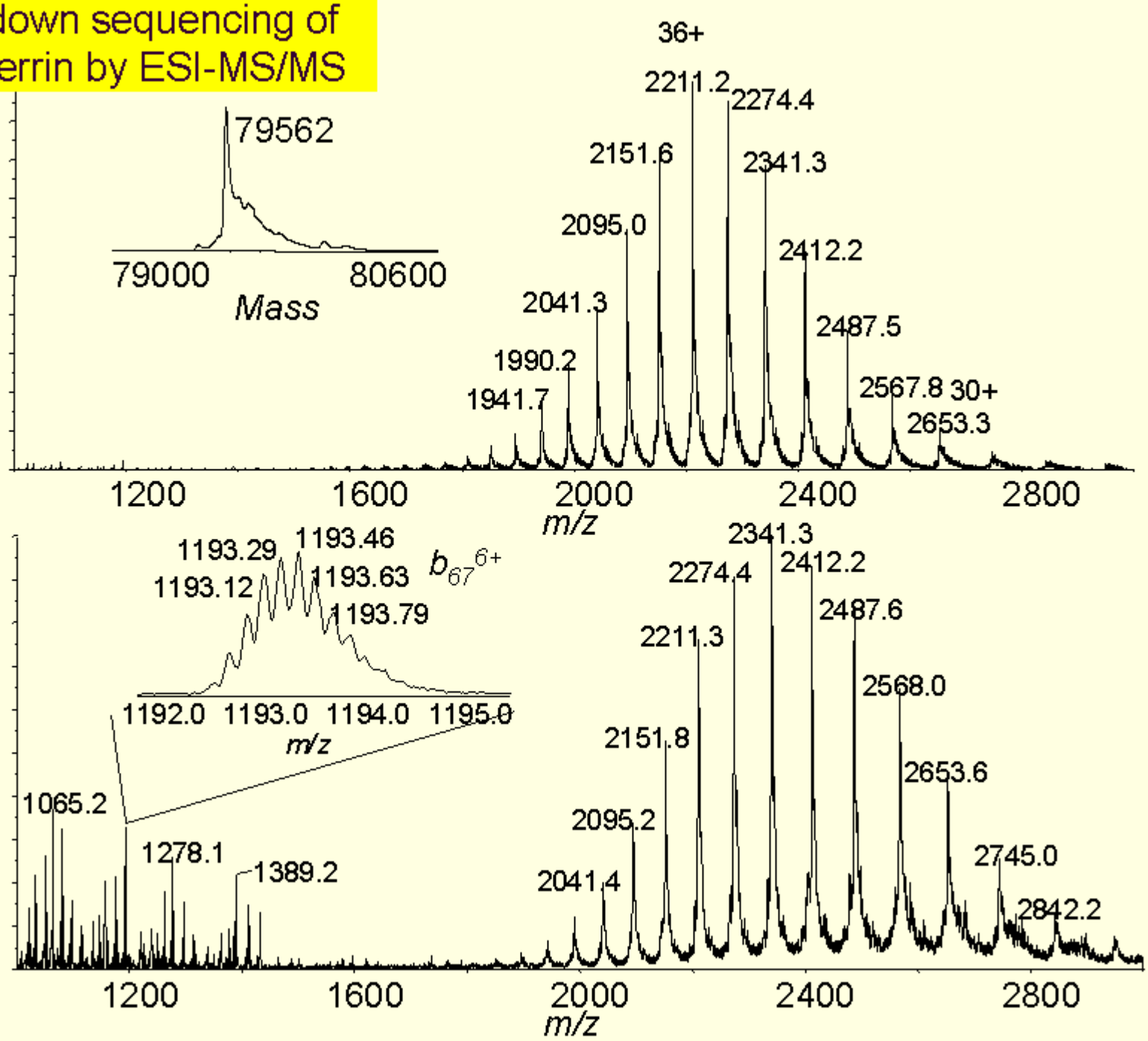
- Often called “nozzle-skimmer dissociation”
- Fragmentation can be controlled by adjustment of nozzle-skimmer voltage
- Low energy CAD



peptide fragmentation in atmosphere / vacuum interface

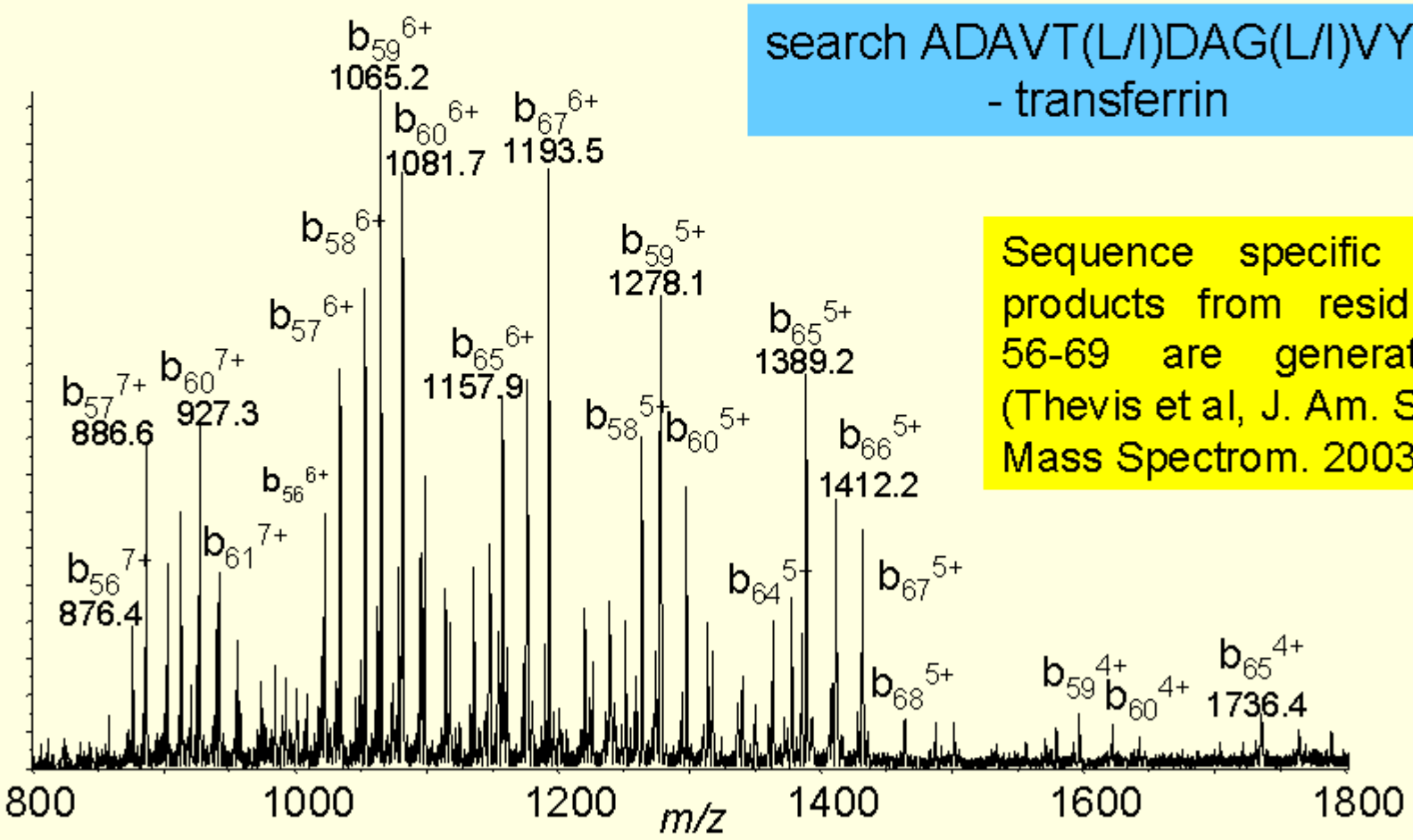
- Useful for obtaining sequence information for relatively pure samples without the need for a tandem mass spectrometer

Top-down sequencing of transferrin by ESI-MS/MS

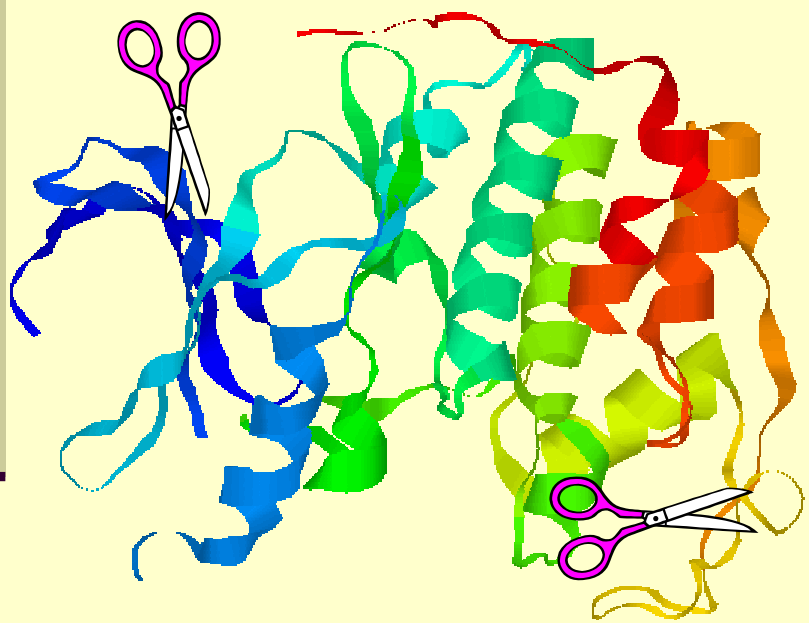


```
1           11          21          31          41
VPDKTVRWCA VSEHEATKCQ SFRDHMKSVI PSDGPSVACV KKASYLDCIR

51          61          71          81
AIAANEADAV TLDAGLVYDA YLAPNNLKPV VAEFYGSKED
```



Mass Spectrometry as a Tool for Protein Crystallography



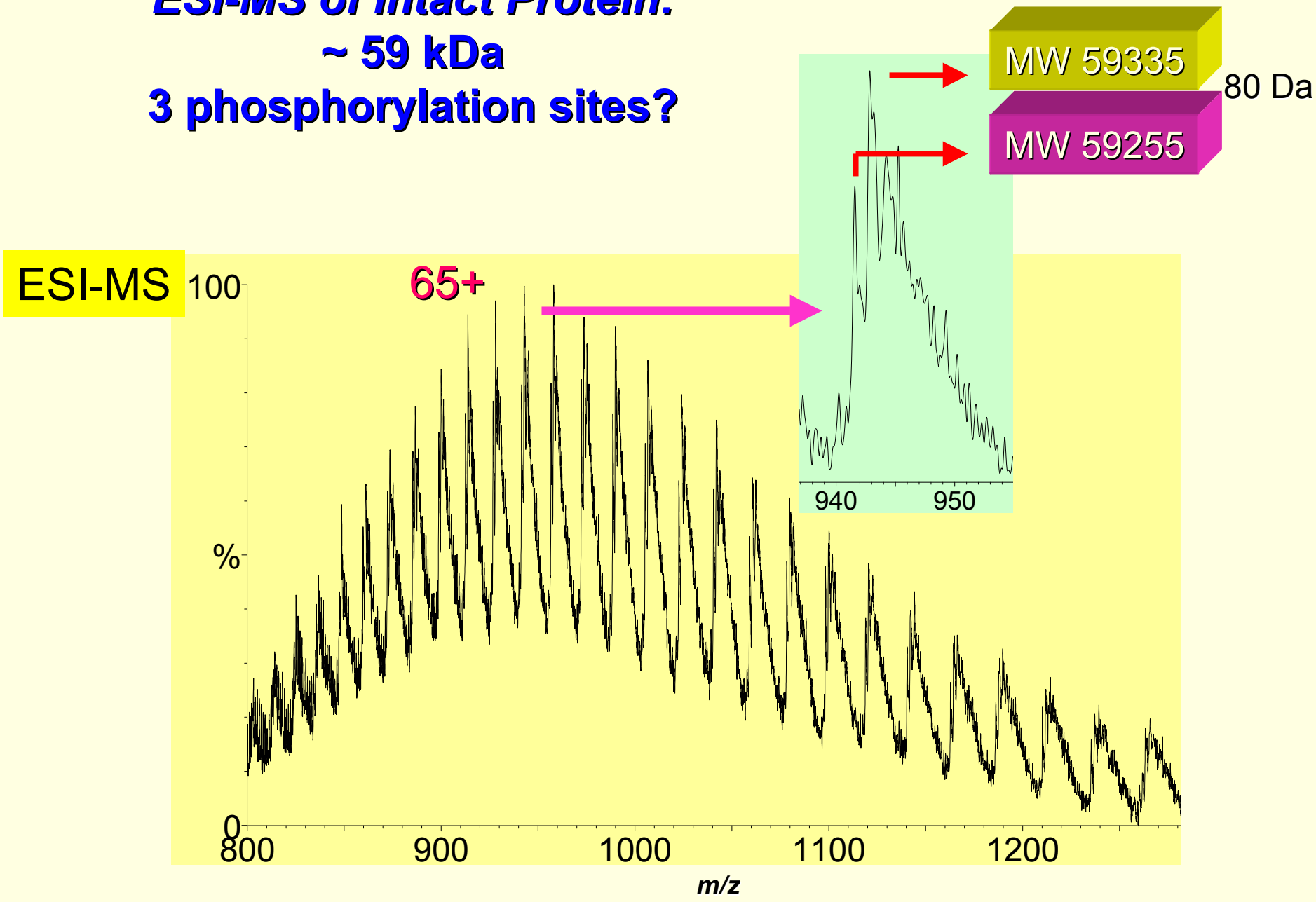
- Construct Verification
- Domain Elucidation
  - Define folding domains
  - Improve chances to generate high quality diffracting crystals
  - ***Limited proteolysis cleaves flexible chains, leaving a more compact folding domain***

(see Cohen & Chait, Annu. Rev. Biophys. Biomol. Struct. 2001)

ESI-MS of Intact Protein:

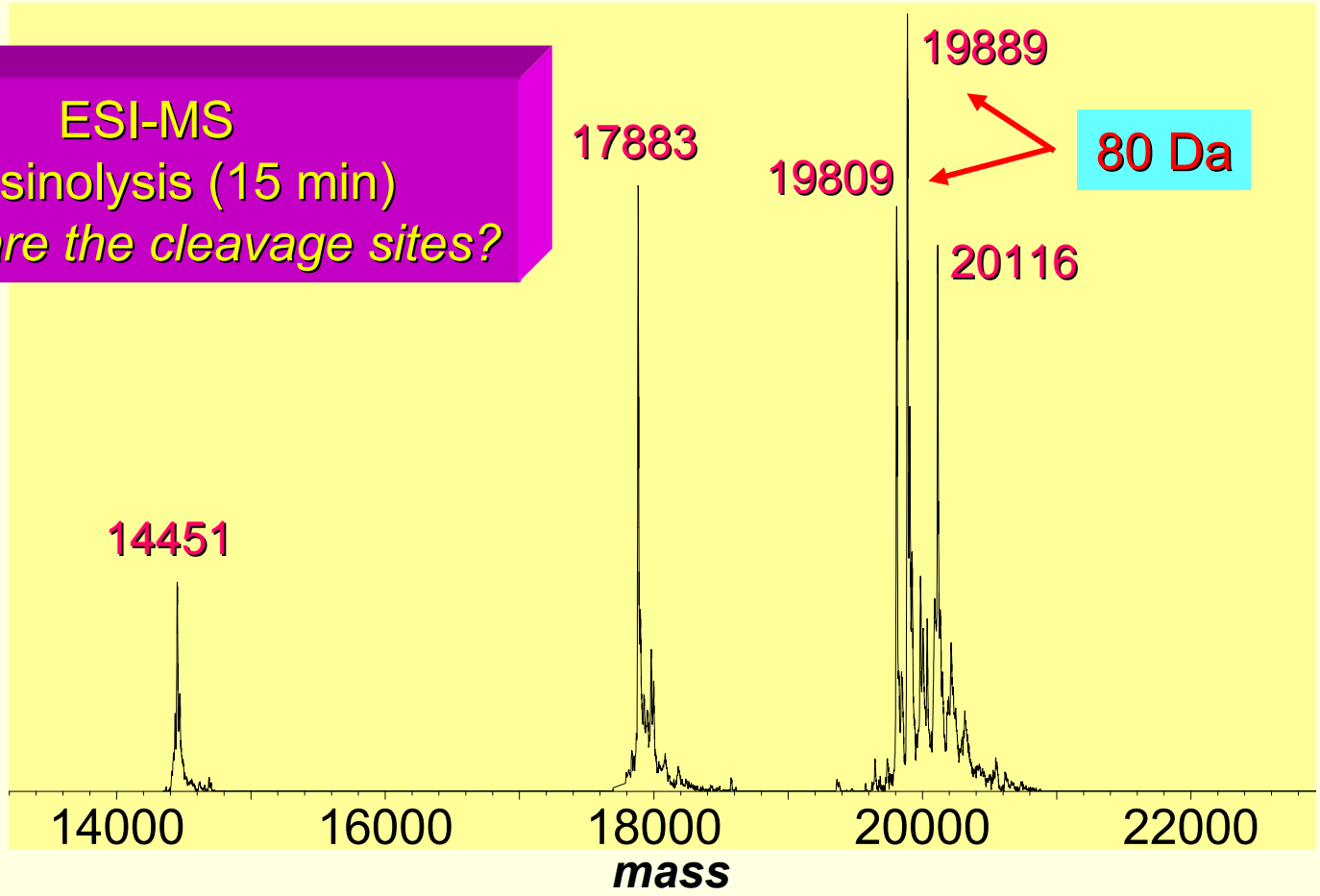
~ 59 kDa

3 phosphorylation sites?



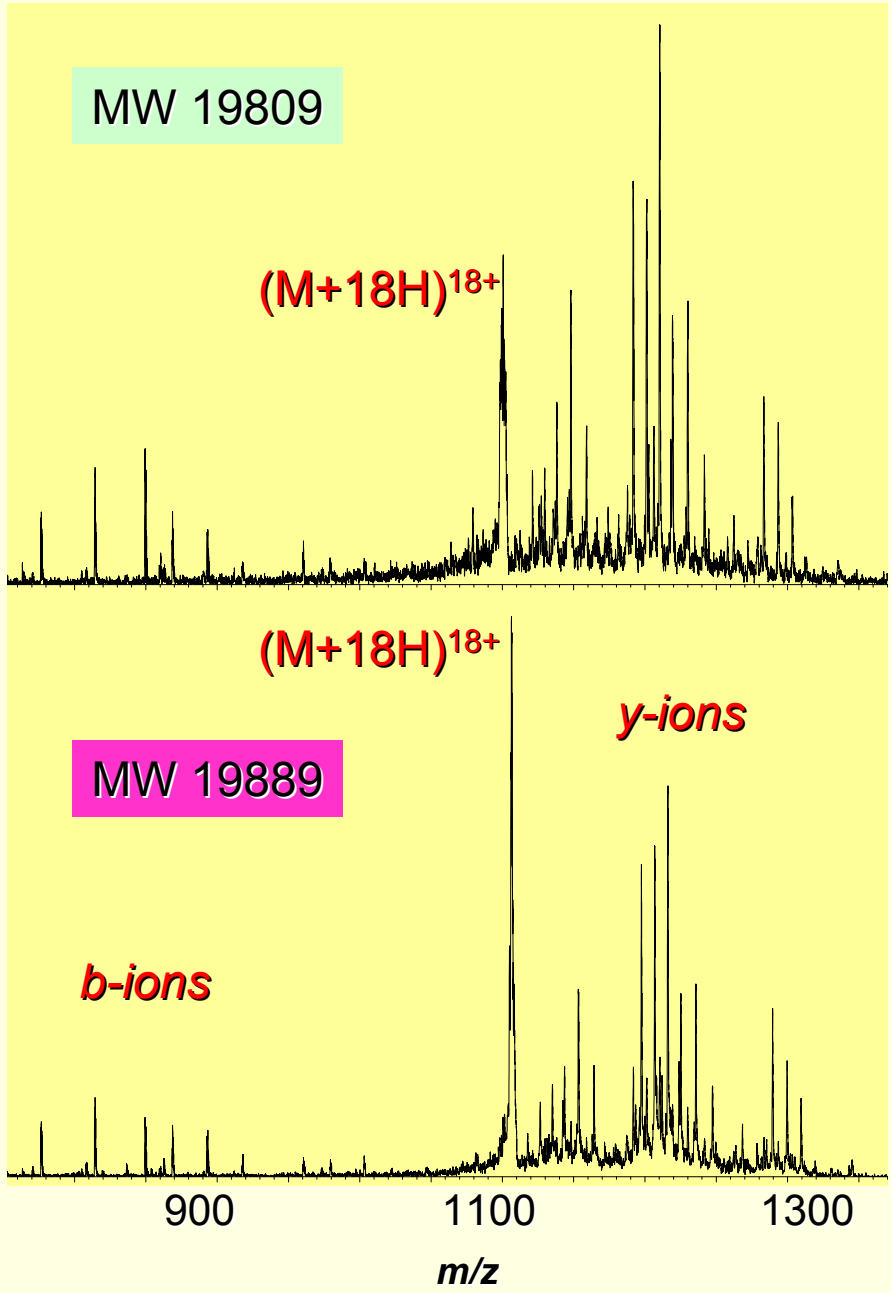
Limited Proteolysis for Domain Elucidation

ESI-MS
Trypsinolysis (15 min)
- *Where are the cleavage sites?*



MS/MS of Intact Proteins

- Complex spectra
- Confirm primary sequence
- Confirm termini, posttranslational modifications



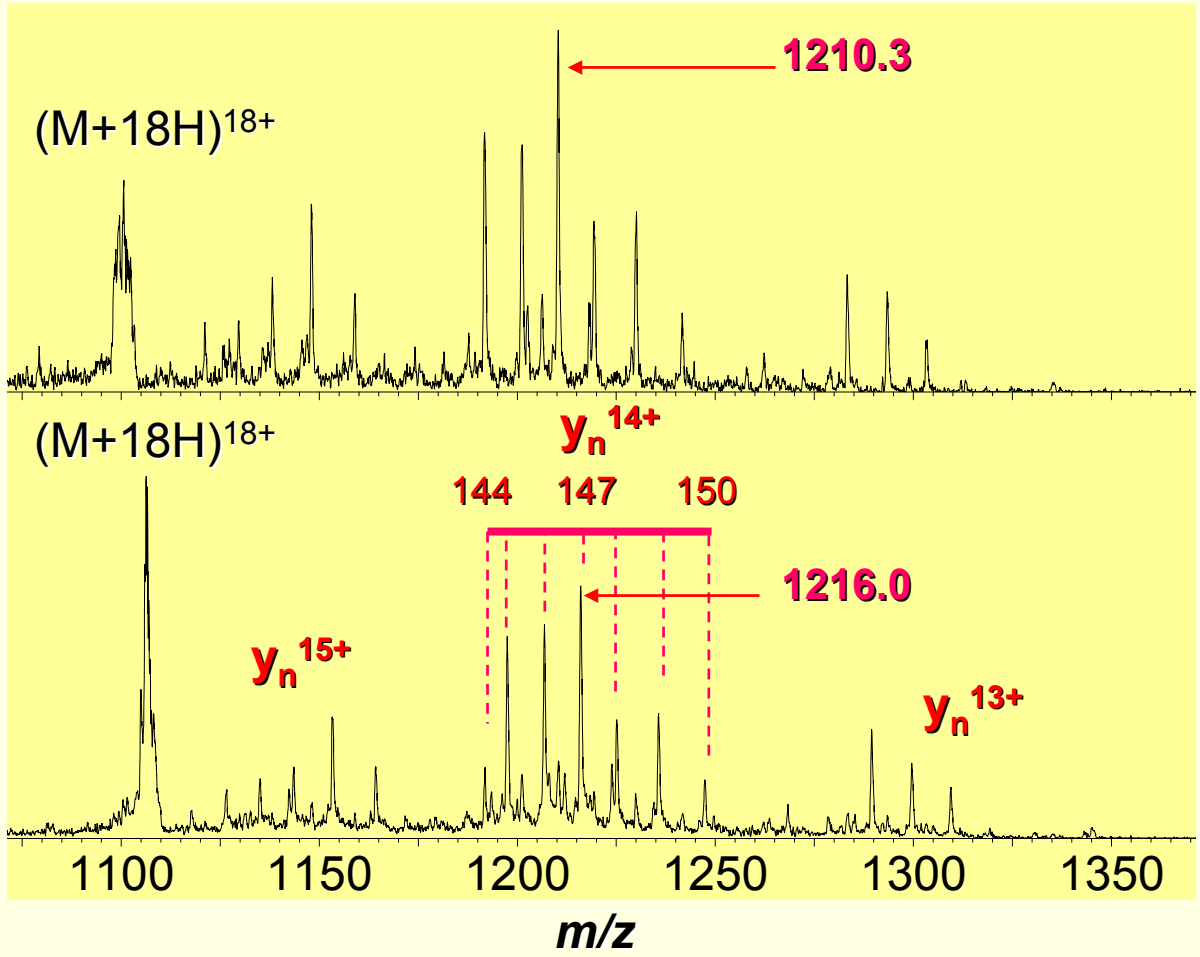
MS/MS of $(M+18H)^{18+}$

y-ions similar, but shifted by 80 Da

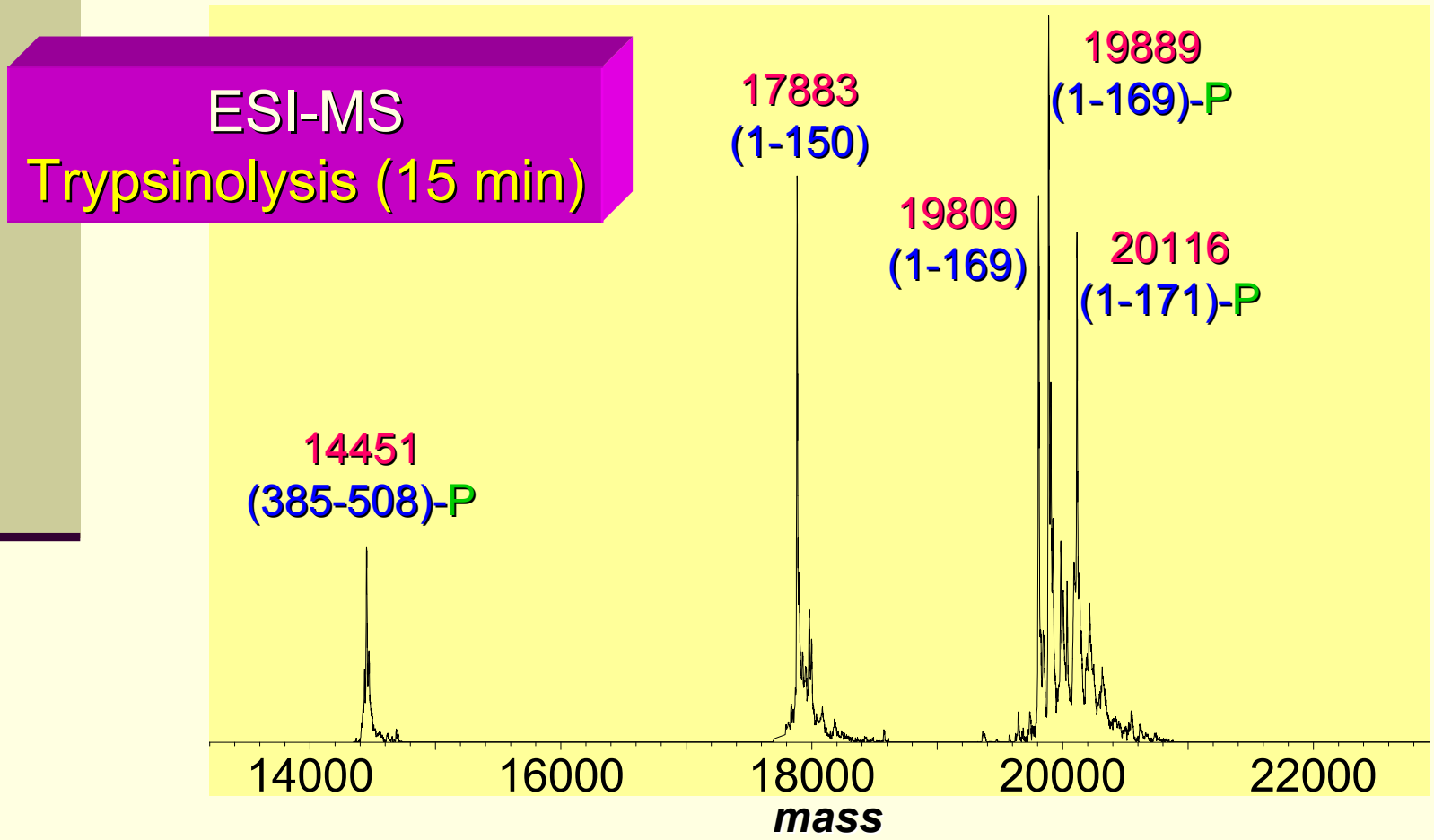
same C-termini, but phosphorylated

MW 19809
residues 1-169

MW 19889
residues 1-169
1 phospho-site



Based on homology modeling and limited proteolysis...



Conclusions

- Bottom-up and top-down approaches for protein characterization are complementary
- Bottom-up methods are robust and lead to efficient processes for identifying proteins
- Top-down methods are not mature, both experimentally and computationally. Protein fragmentation is not predictable (yet). However, information on protein processing and modifications may be flagged by the top-down approach.