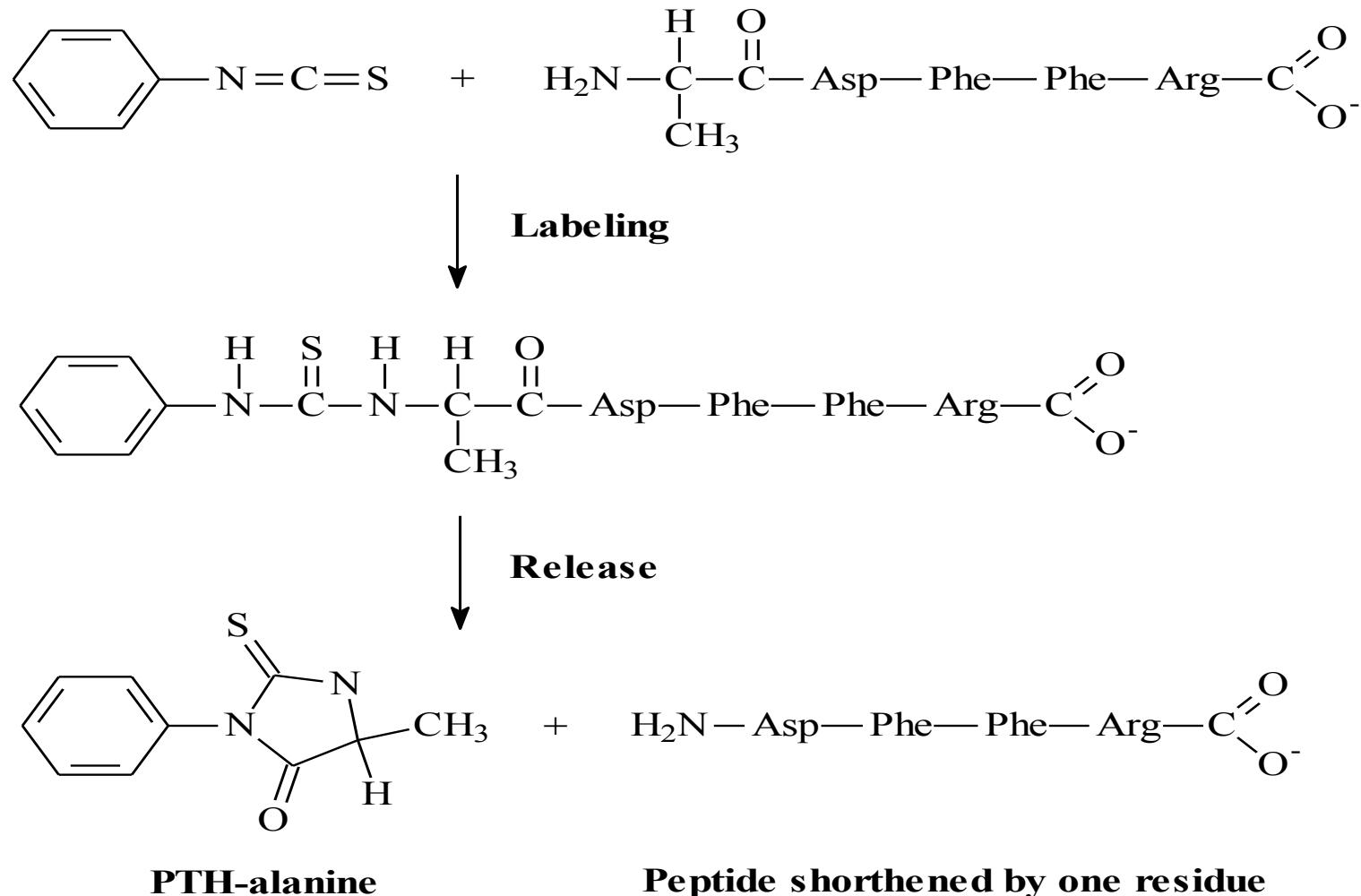


# Peptide Sequencing by Mass Spectrometry

Alex Ramos  
5 April 2005

# Edman degradation

## Phenyl isothiocyanate



# Edman Degradation v. MS/MS

## Protein Identification using Peptide Sequencing

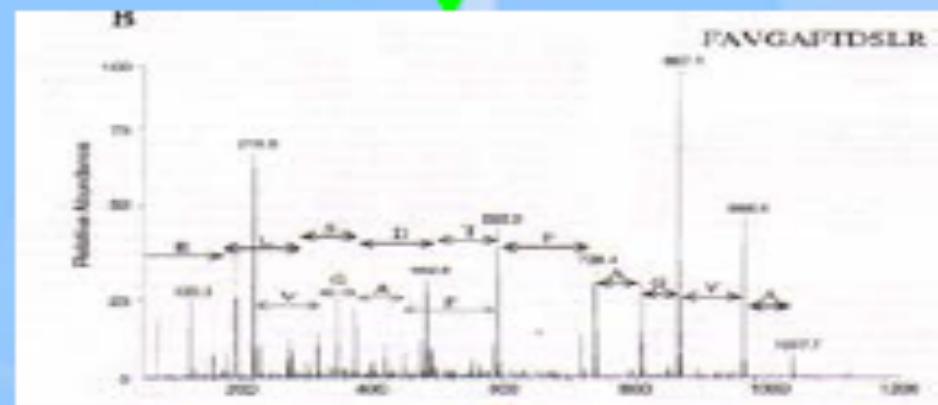
NH<sub>2</sub>-Glu-Gly-Ser-Thr-Ser-Pro-Pro-His-Ala-His-Leu-Lys-COOH

Edman-type degradation

1 hr = Glu  
1 hr = Gly  
1 hr = Ser  
⋮  
1 hr = Lys

Total Time = 12 hours

Tandem mass spectrometry



Total Time = ~1 second

# Why study proteins?

- machines that make cells function
- RNA levels do not always accurately predict protein levels
- targets of drugs

# Peptide Analysis

- Edman Degradation
- MS
  - More sensitive
  - Can fragment peptides faster
  - Does not require proteins or peptides to be purified to homogeneity
  - Has no problem identifying blocked or modified proteins

# Introduction

- MS/MS plays important role in protein identification (fast and sensitive)
- Derivation of peptide sequence an important task in proteomics
- Derivation without help from a protein database (“de novo sequencing”), especially important in identification of unknown protein

# Basic lab experimental steps

1. Proteins digested w/ an enzyme to produce peptides
  2. Peptides charged (ionized) and separated according to their different m/z ratios
  3. Each peptide fragmented into ions and m/z values of fragment ions are measured
- 
- Steps 2 and 3 performed within a tandem mass spectrometer.

# Mass spectrum

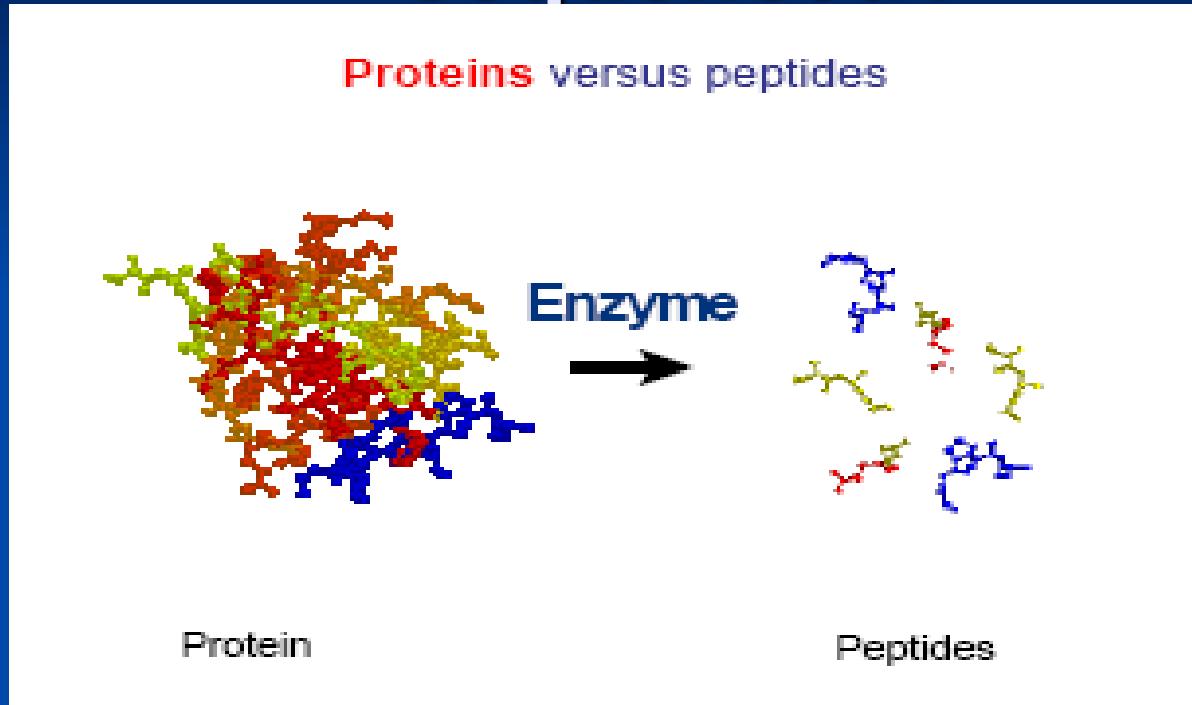
- Proteins consist of 20 different types of a. a. with different masses (except for one pair Leu and Ile)
- Different peptides produce different spectra
- Use the spectrum of a peptide to determine its sequence

# Objectives

- Describe the steps of a typical peptide analysis by MS (proteomic experiment)
- Explain peptide ionization, fragmentation, identification

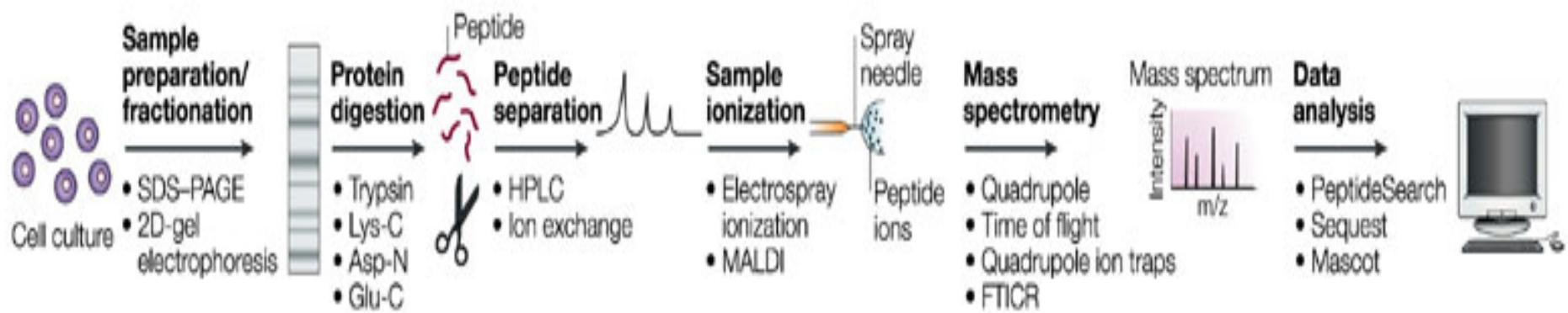
# Why are peptides, and not proteins, sequenced?

---



- Solubility under the same conditions
  - Sensitivity of MS much higher for peptides
  - MS efficiency

# MS Peptide Experiment



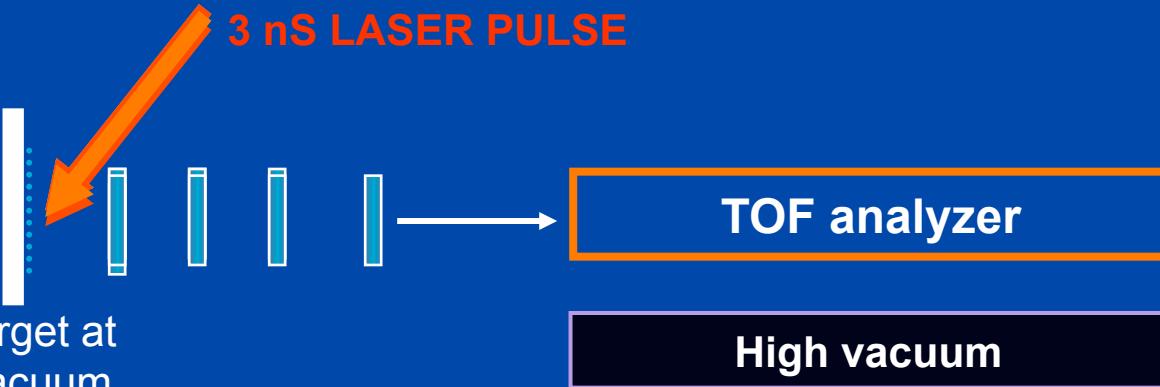
Nature Reviews | Molecular Cell Biology

# Choice of Enzyme

Cleaving agent/ Proteases	Specificity
<b>A. HIGHLY SPECIFIC</b>	
Trypsin	Arg-X, Lys-X
Endoproteinase Glu-C	Glu-X
Endoproteinase Lys-C	Lys-X
Endoproteinase Arg-C	Arg-X
Endoproteinase Asp-N	X-Asp
<b>B. NONSPECIFIC</b>	
Chymotrypsin	Phe-X, Tyr-X, Trp-X, Leu-X
Thermolysin	X-Phe, X-Leu, X-Ile, X-Met, X-Val, X-Ala

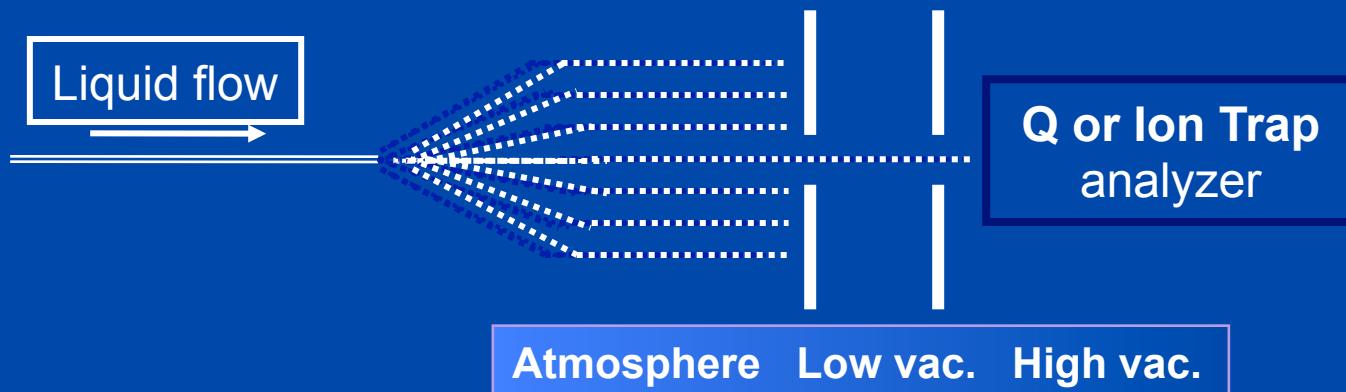
# MALDI

Sample (solid) on target at  
high voltage/ high vacuum

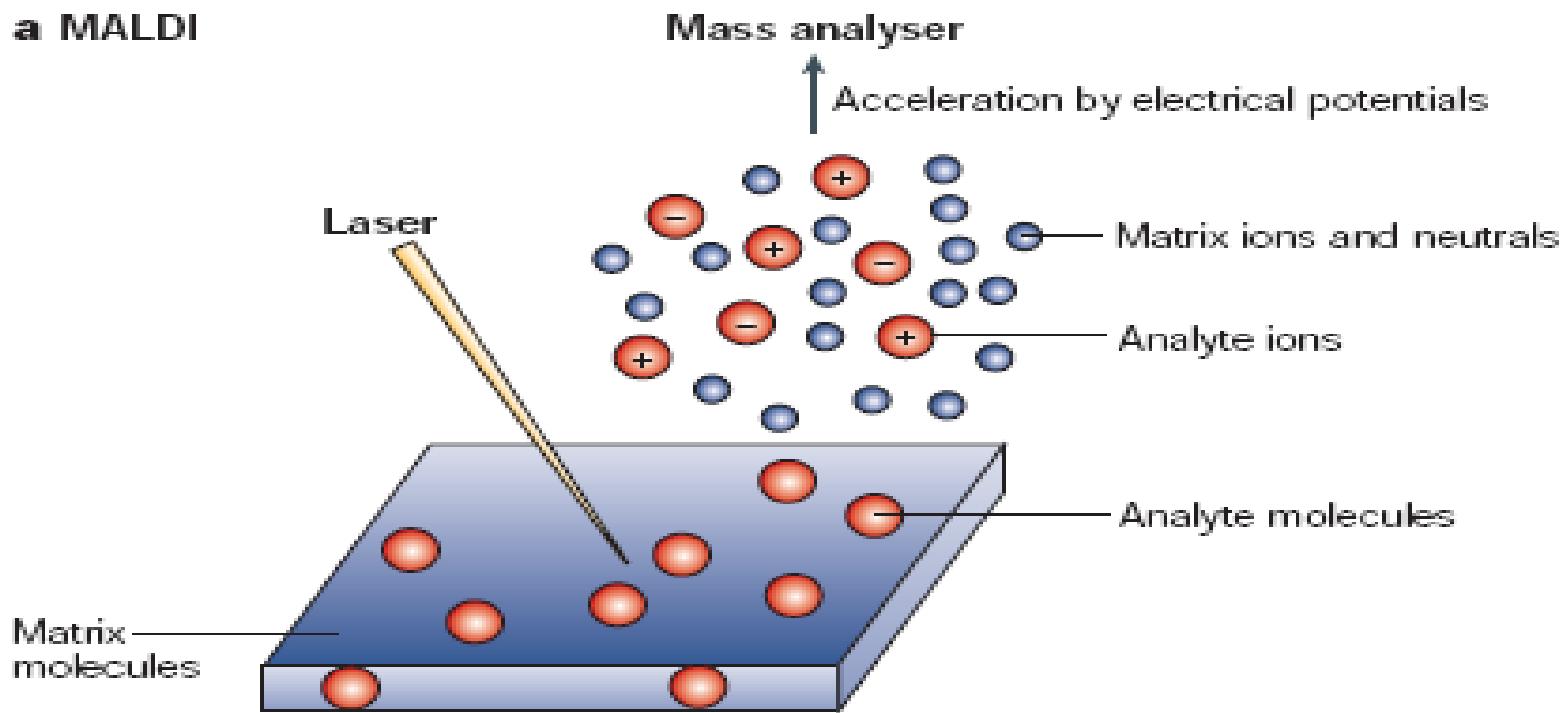
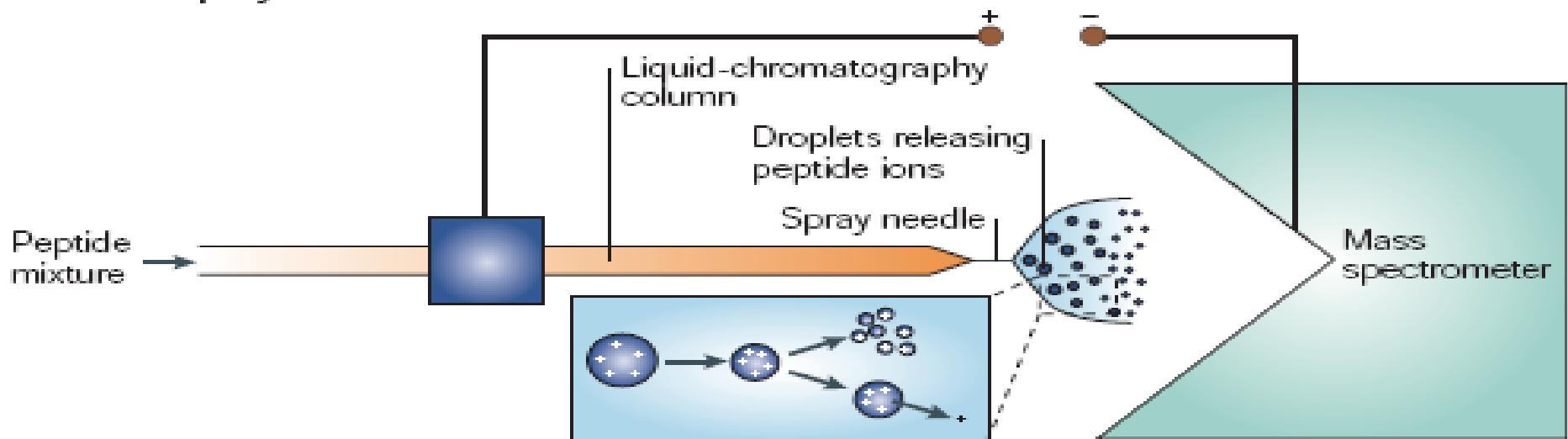


MALDI is a solid-state technique that gives ions in pulses,  
best suited to time-of-flight MS.

# ESI



ESI is a solution technique that gives a continuous stream of ions,  
best for quadrupoles, ion traps, etc.

**a MALDI****b Electrospray ionization**

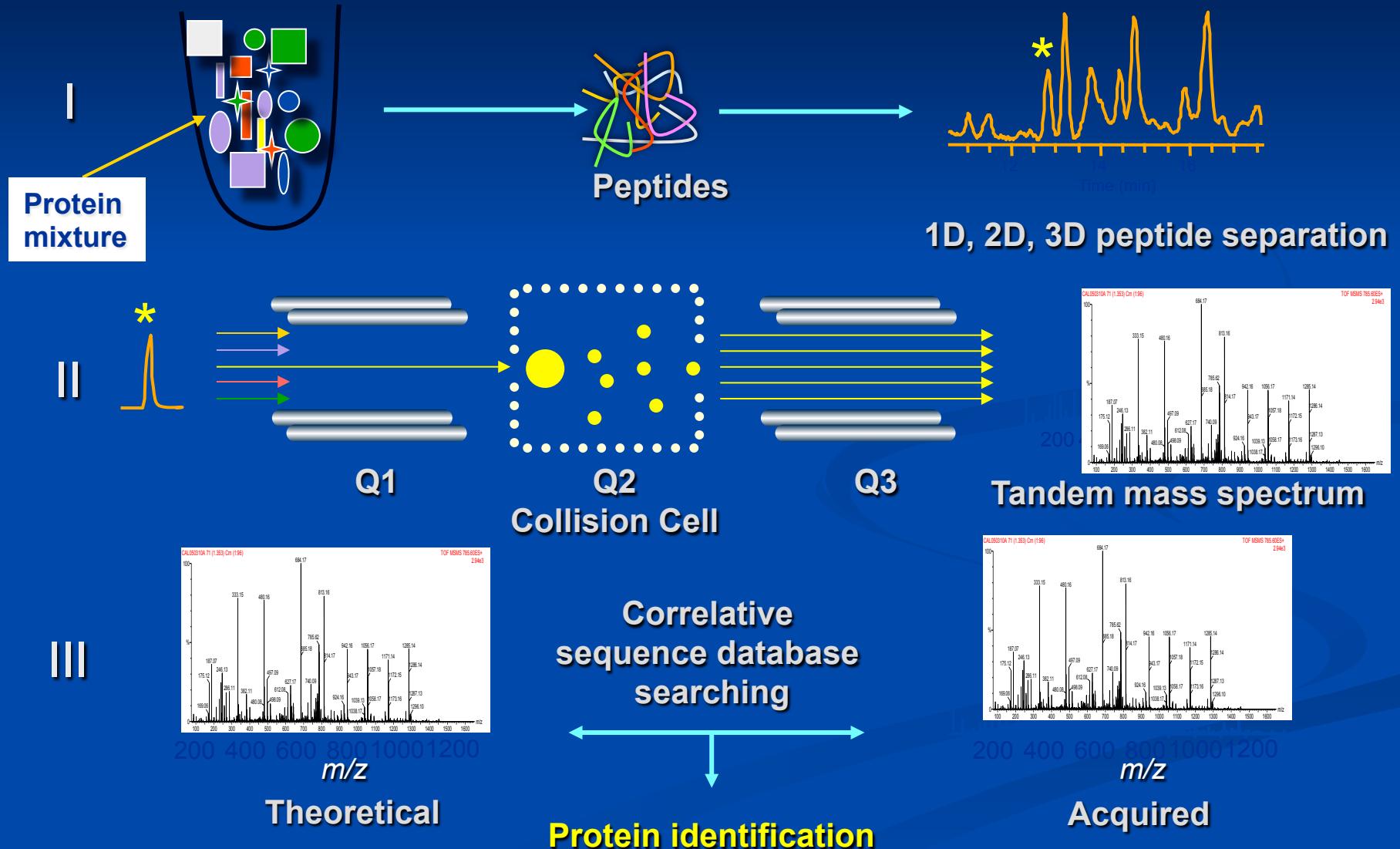
# ....MALDI or Electrospray ?

MALDI is limited to solid state, ESI to liquid

ESI is better for the analysis of complex mixture as it is directly interfaced to a separation techniques (i.e. HPLC or CE)

MALDI is more “flexible” (MW from 200 to 400,000 Da)

# Protein Identification Strategy

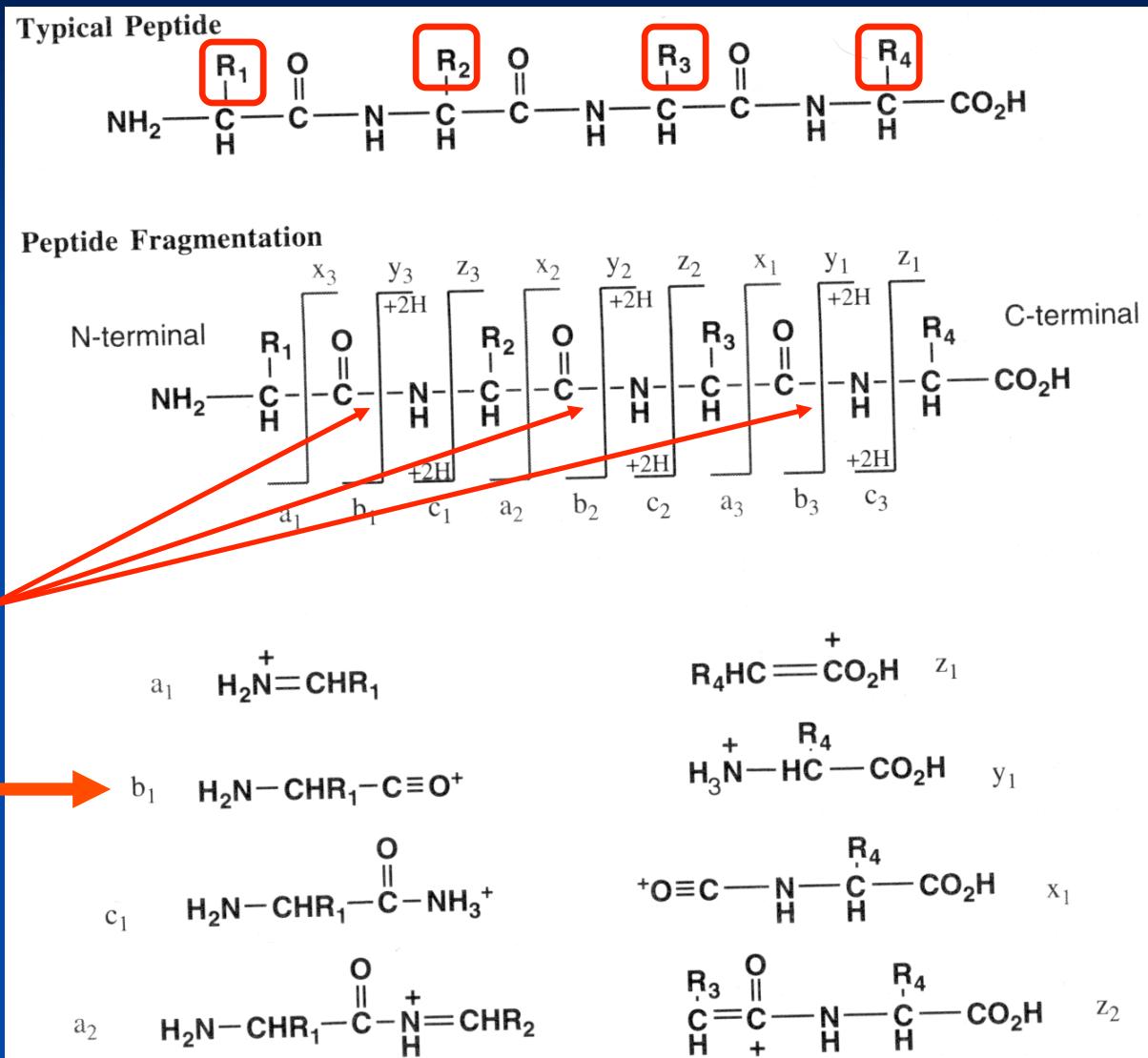


# Breaking Protein into Peptides and Peptides into Fragment Ions

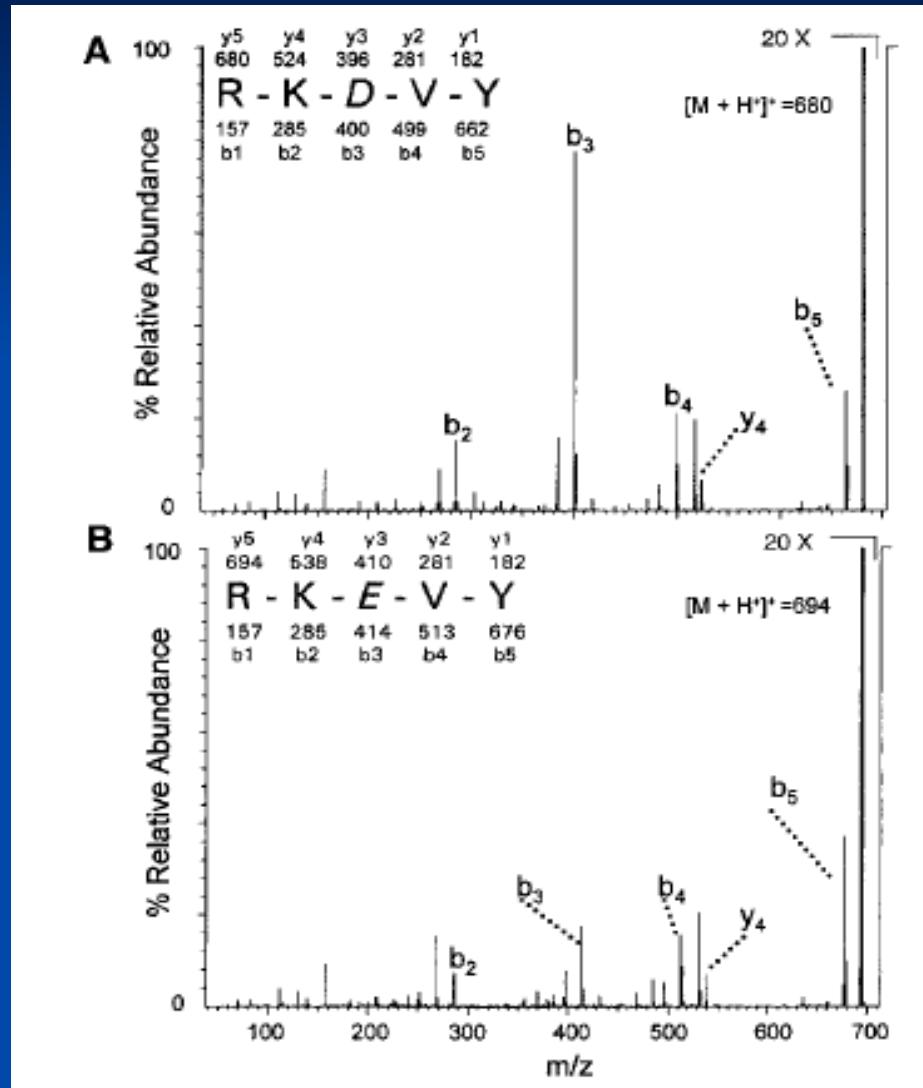
- Proteases, e.g. trypsin, break protein into *peptides*
- MS/MS breaks the peptides down into *fragment ions* and measures the mass of each piece
- MS measure m/z ratio of an ion

# Peptide fragmentation

Amino acids  
differ in their  
side chains



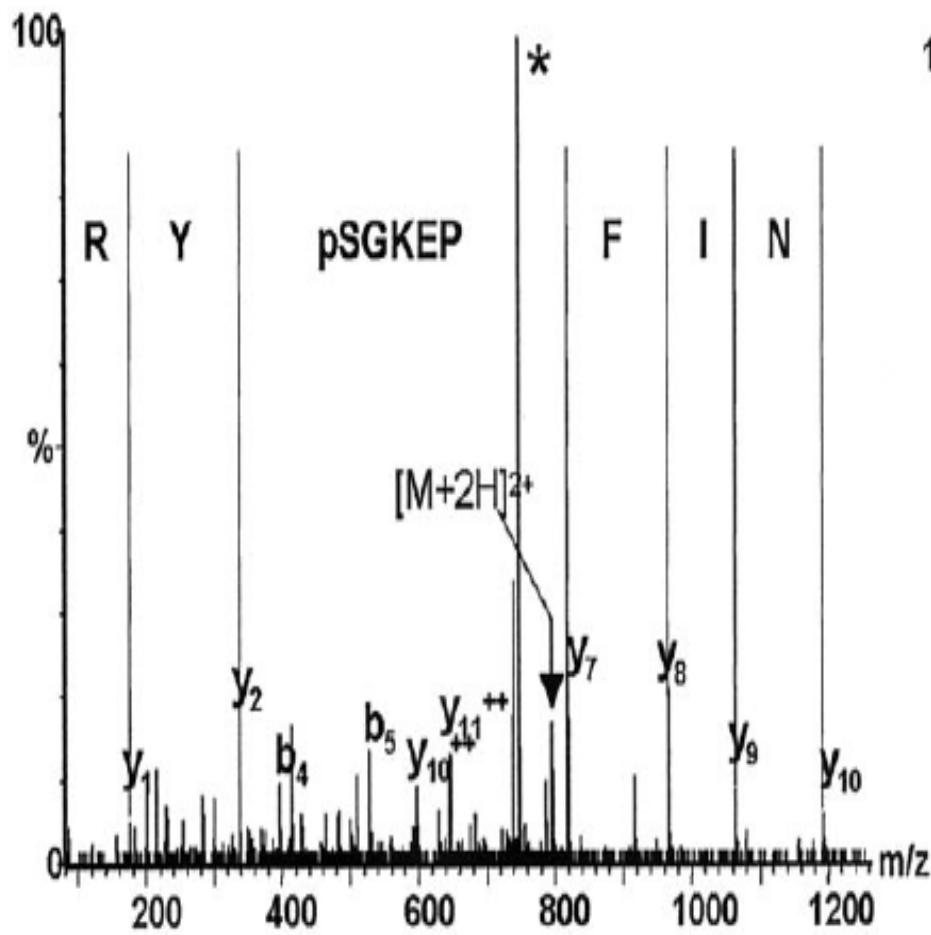
# Tendency of peptides to fragment at Asp (D)



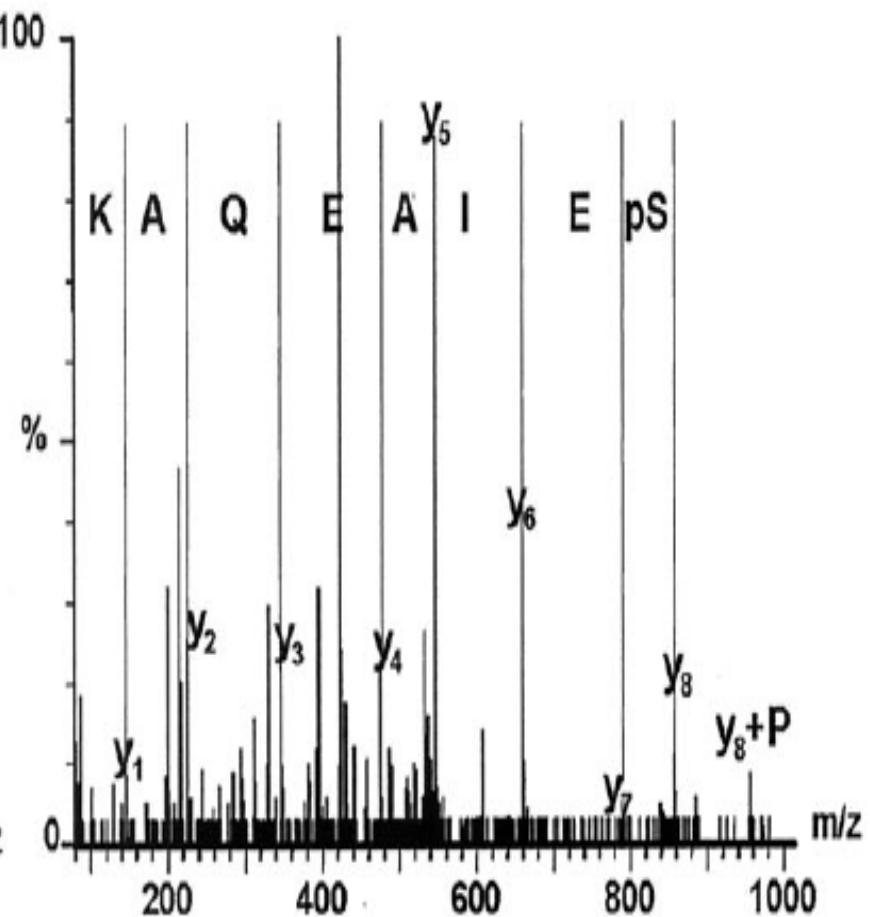
C-terminal side of Asp

Mass Spectrometry in Proteomics  
Ruedi Aebersold\* and David R. Goodlett  
269 Chem. Rev. 2001, 101, 269-295

EAVNIFPEKGpSYR



ELpSEIAEQAK



Large-scale Analysis of *in Vivo* Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry, *Molecular & Cellular Proteomics*, 2003, 2.11, 1234, Thomas S. Nuhse, Allan Stensballe, Ole N. Jensen, and Scott C. Peck

# What you need for peptide mass mapping

- Peptide mass spectrum
- Protein Database
  - GenBank, Swiss-Prot, dbEST, etc.
- Search engines
  - MasCot, Prospector, Sequest, etc.

File Edit View Favorites Tools Help

Back Favorites

Address http://us.expasy.org/sprot/ Go Links

ExPASy Home page Site Map Search ExPASy Contact us PROSITE Proteomics tools

Search Swiss-Prot/TrEMBL for  Go Clear

  
**Swiss-Prot**  
Protein knowledgebase  
**TrEMBL**  
Computer-annotated supplement to Swiss-Prot

The UniProt Knowledgebase consists of

- Swiss-Prot; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [[More details](#) / [References](#) / [Linking to Swiss-Prot](#) / [User manual](#) / [Recent changes](#) / [Disclaimer](#)].
- TrEMBL; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups [at SIB](#) and [at EBI](#).

UniProt Release 4.4 consists of:

Swiss-Prot Release 46.4 of 29-Mar-2005: 178022 entries ([More statistics](#))  
TrEMBL Release 29.4 of 29-Mar-2005: 1647645 entries ([More statistics](#))

> **Swiss-Prot headlines**  
Adding the keyword 'Complete proteome' to fungal entries (Read [more...](#))

Access to Swiss-Prot and TrEMBL

- [SRS](#) - Access to Swiss-Prot, TrEMBL and other databases using the Sequence Retrieval System
- [Full text search](#) in Swiss-Prot and TrEMBL
- [Advanced search](#) in Swiss-Prot and TrEMBL by description, gene name and organism (can be used to create html links to Swiss-Prot/TrEMBL queries)

Internet 2:04

# Database search for protein identification

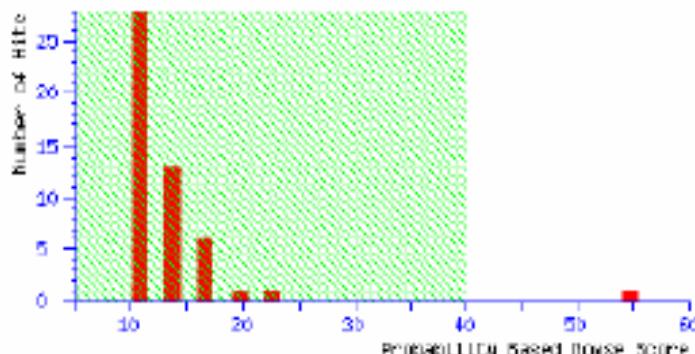
[HOME](#)  
[MASCOT](#)  
[HELP](#)  
[WHAT'S NEW](#)  
[PRODUCTS](#)  
[SUPPORT](#)  
[SITE SEARCH](#)  
[LINKS](#)  
[EMPLOYMENT](#)  
[CONTACT US](#)

## Mascot: Peptide Mass Fingerprint

**Your name**  **Email**   
**Search title**   
**Database** NCBInr  
**Taxonomy** All entries  
**Enzyme** Trypsin  
**Allow up to** 1 missed cleavages  
**Fixed modifications** AB\_old\_ICATd0 (C)  
AB\_old\_ICATd8 (C)  
Acetyl (K)  
Acetyl (N-term)  
Amide (C-term)  
**Variable modifications** AB\_old\_ICATd0 (C)  
AB\_old\_ICATd8 (C)  
Acetyl (K)  
Acetyl (N-term)  
Amide (C-term)  
**Protein mass**  kDa      **Peptide tol. ±** 1.0 Da  
**Mass values**  MH<sup>+</sup>  M<sub>r</sub>      **Monoisotopic**  **Average**   
**Data file**    
**Query**  
NB Contents  
of this field  
are ignored if  
a data file  
is specified.  
**Overview**       **Report top** 20 hits  
   

Copyright © 2000 Matrix Science Ltd. All Rights Reserved. Last Updated 03/06/2003 12:50:29

# Peptide sequencing using MASCOT



## Peptide Summary Report

[Switch to Protein Summary Report](#)

To create a bookmark for this report, right click this link: [Peptide Summary Report \(./data/20021008/FotcIca.dat\)](#)

Select All

Select None

Search Selected

Error tolerant

1. [gi|16924319](#) Mass: 40477 Total score: 55 Peptides matched: 1

(BC017450) Unknown (protein for IMAGE:3538275) [Homo sapiens]

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
<input checked="" type="checkbox"/>	<a href="#">14</a>	895.70	1789.39	1789.88	-0.50	0	55	<a href="#">SYELPDGQVITIGNER</a>

Proteins matching the same set of peptides:

[gi|4501887](#) Mass: 41766 Total score: 55 Peptides matched: 1

(NM\_001614) actin, gamma 1 propeptide; cytoskeletal gamma-actin; actin, cytoplasmic 2 [Homo sapiens]

[gi|16359158](#) Mass: 41736 Total score: 55 Peptides matched: 1

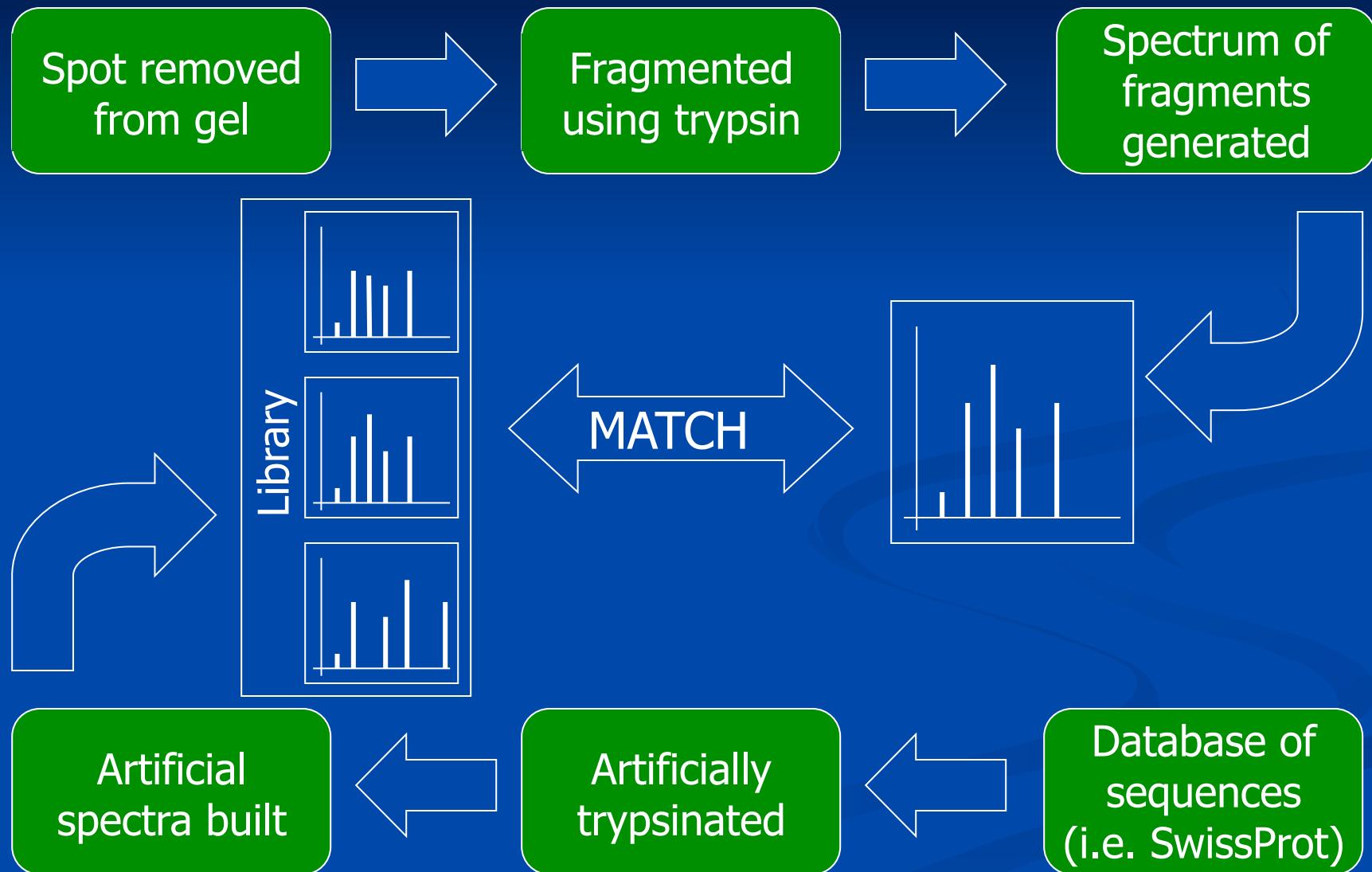
(BC016045) actin, beta [Homo sapiens]

[gi|4885049](#) Mass: 41992 Total score: 55 Peptides matched: 1

(NM\_005159) actin, alpha, cardiac muscle precursor [Homo sapiens]

[gi|14714562](#) Mass: 18762 Total score: 55 Peptides matched: 1

# Protein Identification by MS



# Conclusions

- MS of peptides enables high throughput identification and characterization of proteins in biological systems
- “de novo sequencing” can be used to identify unknown proteins not found in protein databases

# References

H. Steen and M. Mann. “The ABC’ s (and XYZ’ s) of Peptide Sequencing” *Molecular Cell Biology, Nature Reviews.* 2004, 5, 699.

T. S. Nuhse, A. Stensballe, O. Jensen, and S. Peck. “Large-scale Analysis of *in Vivo* Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry” *Molecular & Cellular Proteomics*, 2003, 2.11, 1234.

R. Aebersold and D. Goodlett. “Mass Spectrometry in Proteomics” *Chem. Rev.*, 2001, 101, 269.