

# **Lecture 1**

## **Introduction- Protein Sequencing**

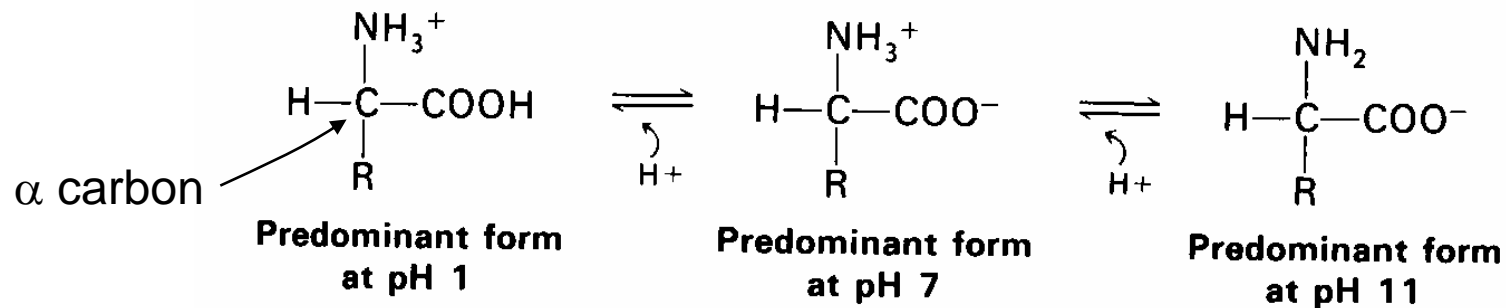
### **Production of Ions for Mass Spectrometry**

Nancy Allbritton, M.D., Ph.D.  
Department of Physiology & Biophysics  
824-9137 (office)  
nlallbri@uci.edu

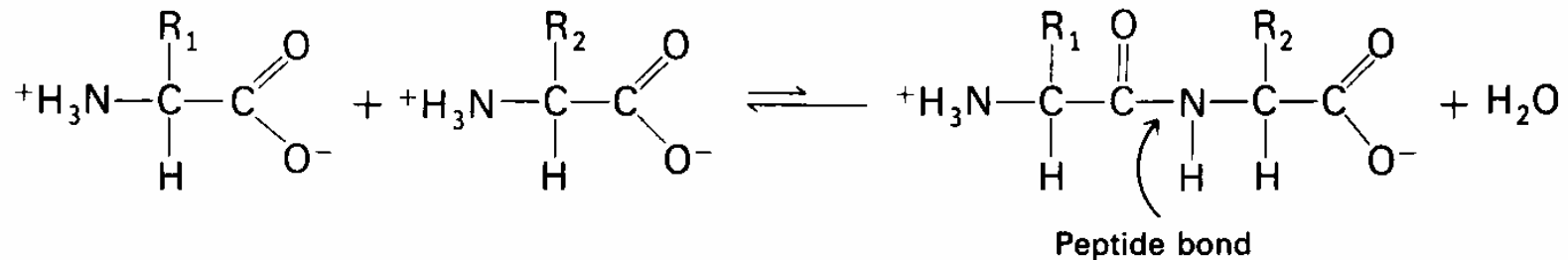
Office- Rm D349  
Medical Science D Bldg.

# Introduction to Proteins

Amino Acid- structural unit of a protein



Amino acids- linked by peptide (amide) bond

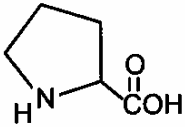
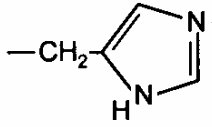
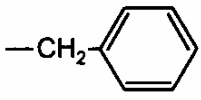
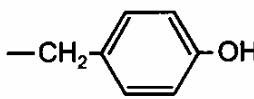
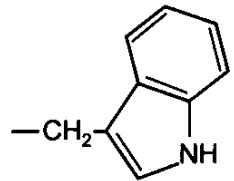


# Amino Acids

Proteins- 20 amino acids  
(Recall DNA- 4 bases)

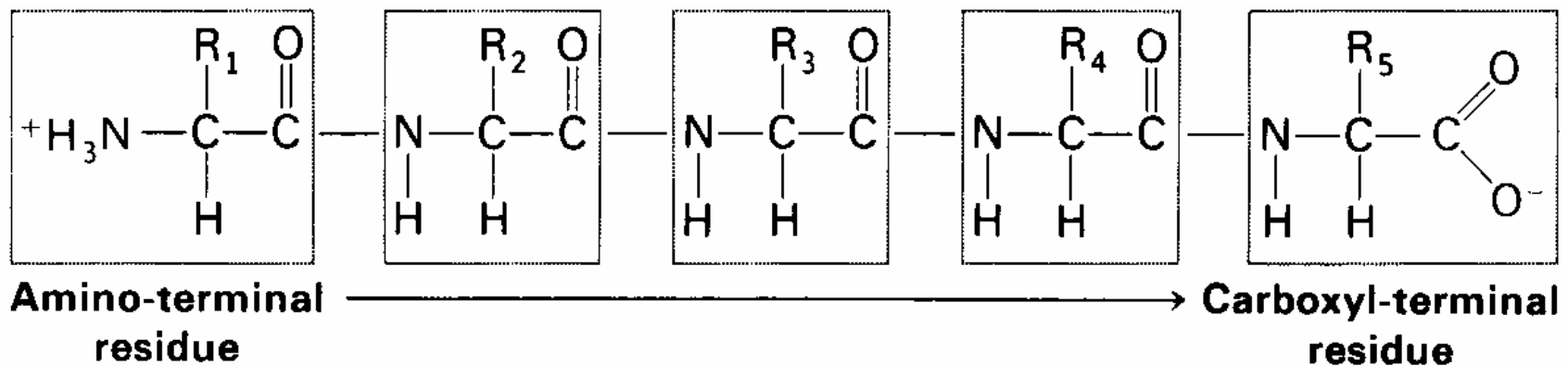
## R groups-

Varying size,  
shape, charge,  
H-bonding capacity,  
& chemical reactivity

Glycine, G	$\text{—H}$	Aspartate, D	$\text{—CH}_2\overset{\text{O}}{\parallel}\text{COH}$
Alanine, A	$\text{—CH}_3$	Glutamine, Q	$\text{—CH}_2\text{CH}_2\overset{\text{O}}{\parallel}\text{CNH}_2$
Serine, S	$\text{—CH}_2\text{OH}$	Glutamate, E	$\text{—CH}_2\text{CH}_2\overset{\text{O}}{\parallel}\text{COH}$
Proline, P		Lysine, K	$\text{—CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2$
Valine, V	$\text{—}\underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{C}}}\text{H—}$	Methionine, M	$\text{—CH}_2\text{CH}_2\text{SCH}_3$
Threonine, T	$\text{—}\underset{\text{CH}_3}{\overset{\text{OH}}{\text{C}}}\text{H—}$	Histidine, H	
Cysteine, C	$\text{—CH}_2\text{SH}$	Phenylalanine, F	
Leucine, L	$\text{—CH}_2\underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{C}}}\text{H—}$	Arginine, R	$\text{—CH}_2\text{CH}_2\text{CH}_2\text{NHC}(\text{NH}_2)=\text{NH}$
Isoleucine, I	$\text{—}\underset{\text{CH}_3}{\text{CH}}\text{CH}_2\text{CH}_3$	Tyrosine, Y	
Asparagine, N	$\text{—CH}_2\overset{\text{O}}{\parallel}\text{CNH}_2$	Tryptophan, W	

# Introduction to Proteins

Polypeptide Chain (Protein) - Many amino acids linked by peptide bonds



By convention: Residue 1 starts at amino terminus.

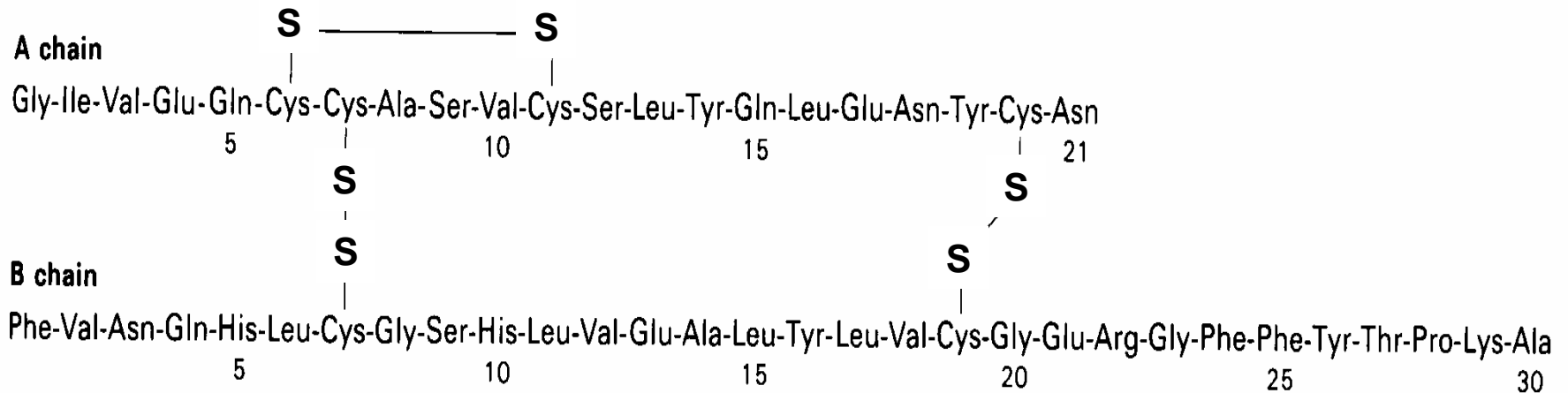
- Polypeptides-
- a. Main chain i.e. regularly repeating portion
  - b. Side chains- variable portion

# Introduction to Proteins

25,000 human genes  $\longrightarrow$   $\geq 2 \times 10^6$  proteins

Natural Proteins - Typically 50-2000 amino acids  
i.e. 550-220,000 molecular weight

Over 200 different types of post-translational modifications.  
Ex: proteolysis, phosphorylation, acetylation, glycosylation



Ex: Insulin

# Protein Complexity Is Very Large

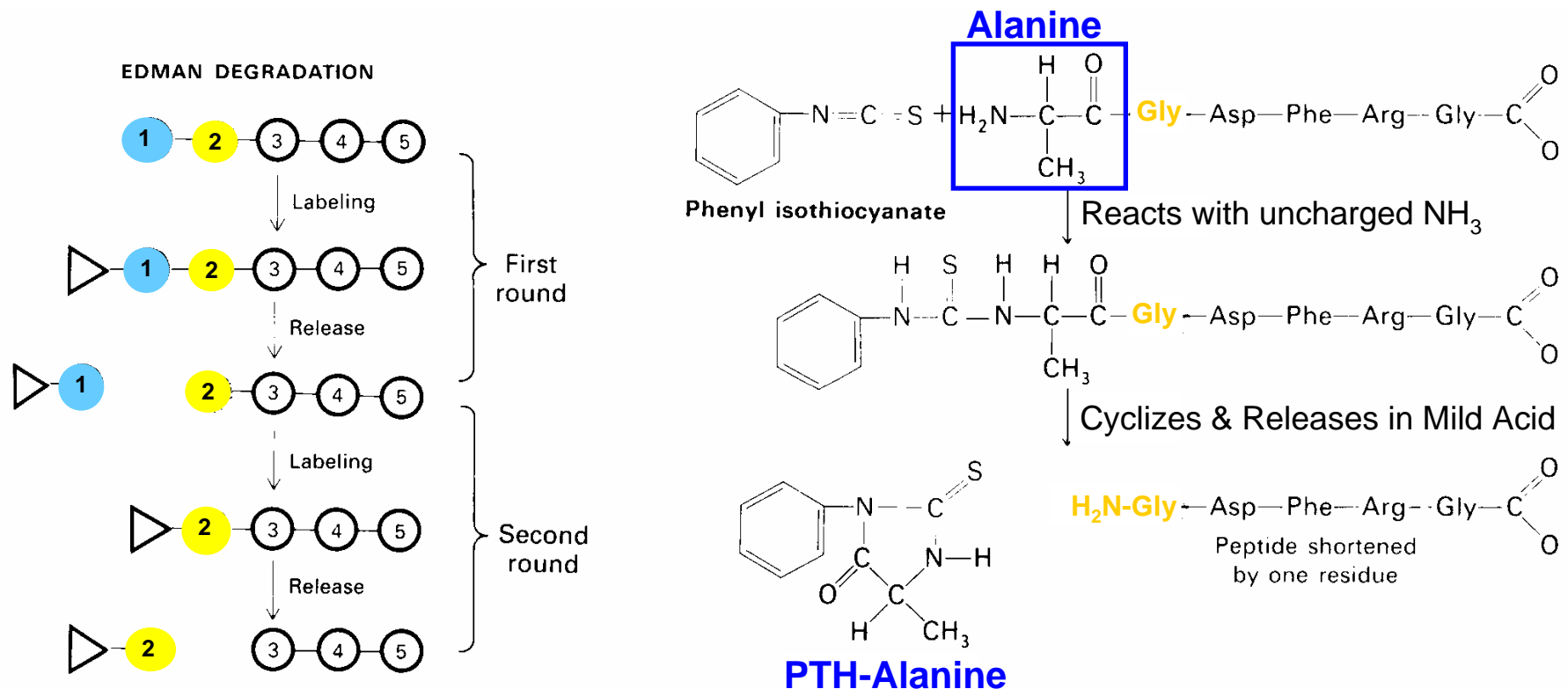
Over 200 different types  
of post-translational  
modifications.

**Table 7.1** Post-translational modifications and corresponding mass variations

Post-translational modification	Mass difference (Da)
Methylation	14.03
Propylation	42.08
Sulfation	80.06
Phosphorylation	79.98
Glycosylations by:	
Deoxyhexoses (Fuc)	146.14
Hexosamines (GlcN, GalN)	161.16
Hexoses (Glc, Gal, Man)	162.14
<i>N</i> -Acetylhexosamines (GlcNAc, GalNAc)	203.19
Pentoses (Xyl, Ara)	132.12
Sialic acid (NeuNAc)	291.26
Reduction of a disulfide bridge	2.02
Carbamidomethylation	57.03
Carboxymethylation	58.04
Cysteinylation	119.14
Ethylpyridylation	105.12
Acetylation	42.04
Formylation	28.01
Biotinylation	226.29
Farnesylation	204.36
Myristoylation	210.36
Pyridoxal phosphate Schiff condensation	231.14
Stearoylation	266.47
Palmitoylation	238.41
Lipoylation	188.30
Carboxylation of Asp or Glu	44.01
Deamidation of Asn or Gln	0.98
Hydroxylation	16.00
Methionine oxidation	16.00
Proteolysis of a peptide bond	18.02
Deamination from Gln to pyroglutamic	-17.03

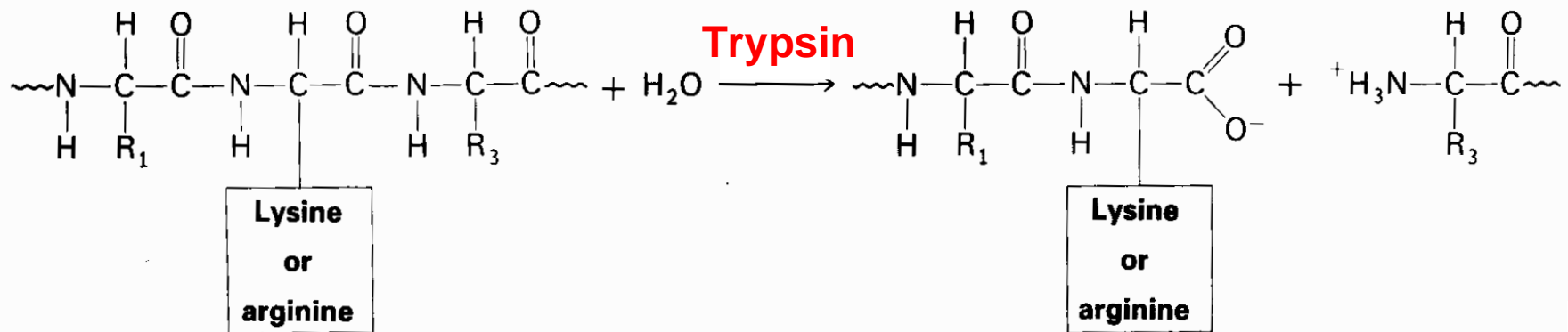
# The Problem of Protein Sequencing.

**Edman Degradation:** Step-wise cleavage of an amino acid from the amino terminus of a peptide.



# Edman Degradation

1. Must be short peptide ( $\leq 50$  a.a.)  
amino acid release- 98% efficiency  
proteins- must fragment (CNBr or trypsin)



2. Frequently fails due to a blocked amino terminus
3. Intolerant of impurities
4. Tedious & time consuming (hours-days)  
1 amino acid cycle- 2 hours



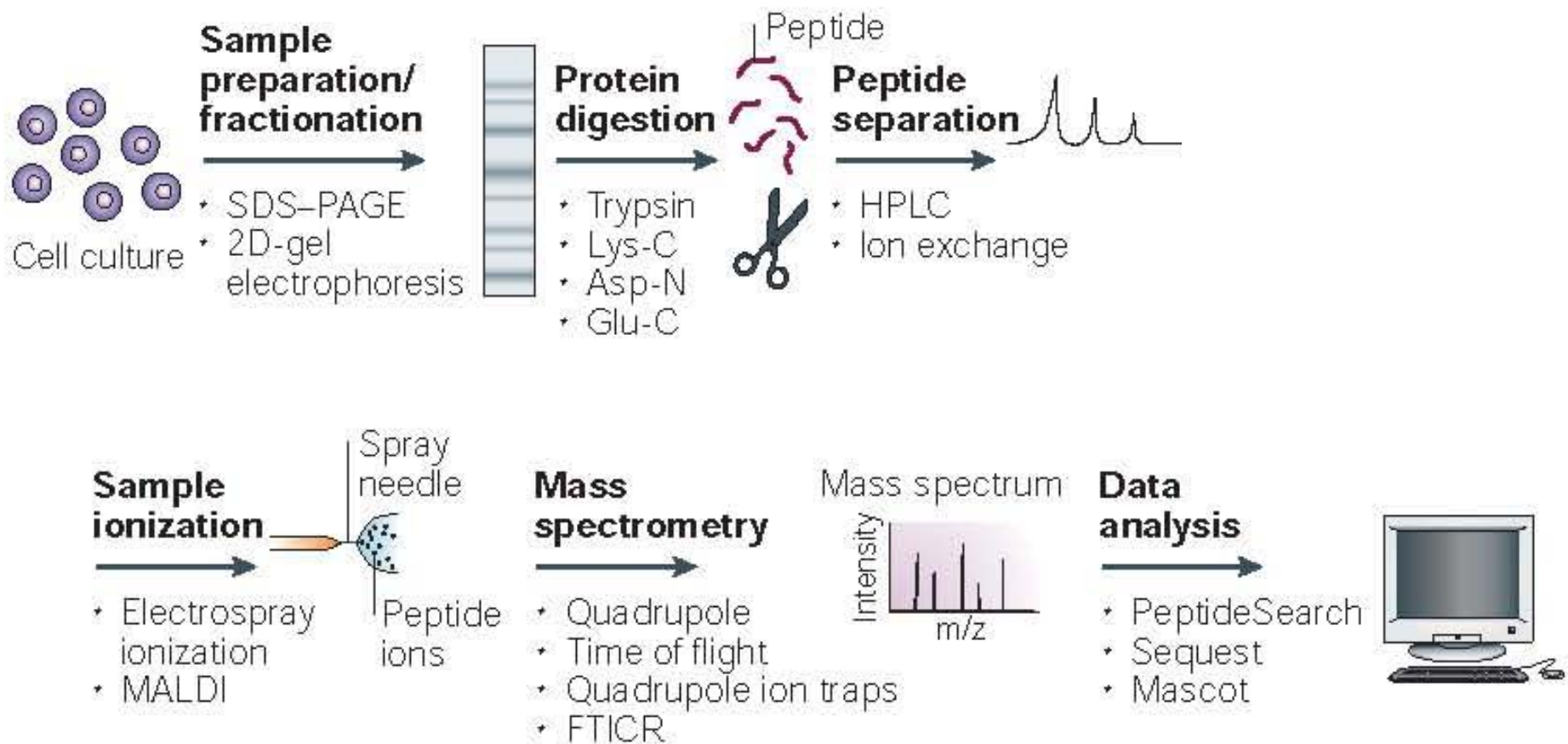
## **Solution: Mass Spectrometry**

1. Impurities are tolerated.
2. Fast (sec - min).
3. Amino terminus can be blocked.
4. Always get some information.

## **Steps in Mass Spectrometry**

1. Production of Ions
2. Ion Separation
3. Ion Detection
4. Data Acquisition & Reduction

# Protein Sequencing By MS



Steen & Mann Nat, Rev, Molec. Cell Bio. 2004, 5:699-711.

# Units and Numbers

$m/z$  = mass/(# of e charges on the ion)

$m$  = Daltons (Da) or atomic mass units (U)

$m/z$  units = Thompson (Th)

Ex: Methyl Chloride ( $\text{CH}_3\text{Cl}$ )

$m_{\text{H}} = 1$ ;  $m_{\text{Cl}} = 35$  (76%);  $m_{\text{Cl}} = 37$  (24%)

$m_{\text{C}} = 12$  (99%) &  $m_{\text{C}} = 13$  (1%)

Atomic Weight = average mass using isotope prevalence

50.5 Da

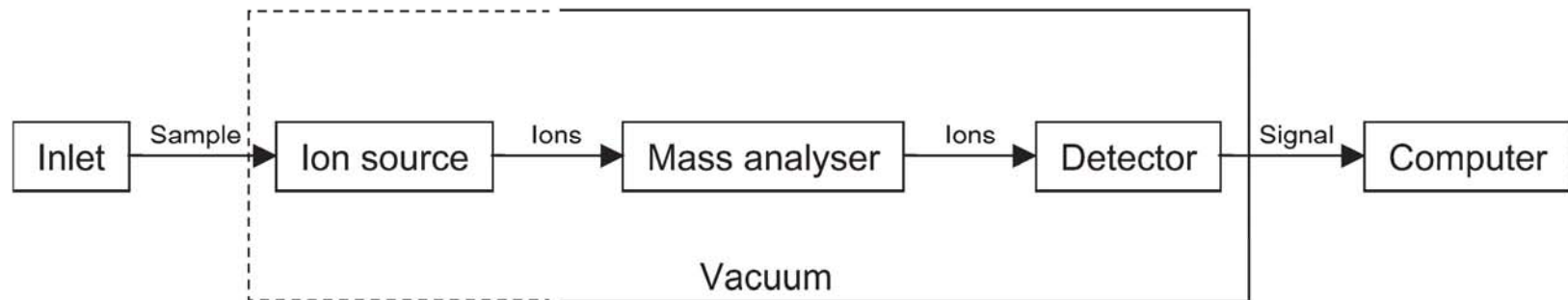
MS gives peaks for monoisotopic masses i.e.

Two major peaks- 50 & 52 Da

Very minor peaks- 51 & 53 Da

# Elements of a Mass Spectrometer

1. Device to insert sample into the mass spec.-  
sample probe, chromatograph,  
capillary
2. Source to produce ions from the sample.
3. Analyzer ( $\geq 1$ ) to separate ions by  $m/z$ .
4. Detector to count ions.
5. Computer to control instrument and  
collect & analyze data.



## Mean Free Path of an Ion

L = mean free path traveled by a molecule before colliding with another

$$L = kT/[p\sigma(2)^{1/2}] \quad ; \quad k = \text{Boltzman Constant}$$

T = temperature

p = pressure

$\sigma$  = collision X-section =  $\pi d^2$

d = sum of radii of  
stationary & colliding ions

$$L \text{ (cm)} = 4.95/p \quad \text{for } p \text{ in milliTorr}$$

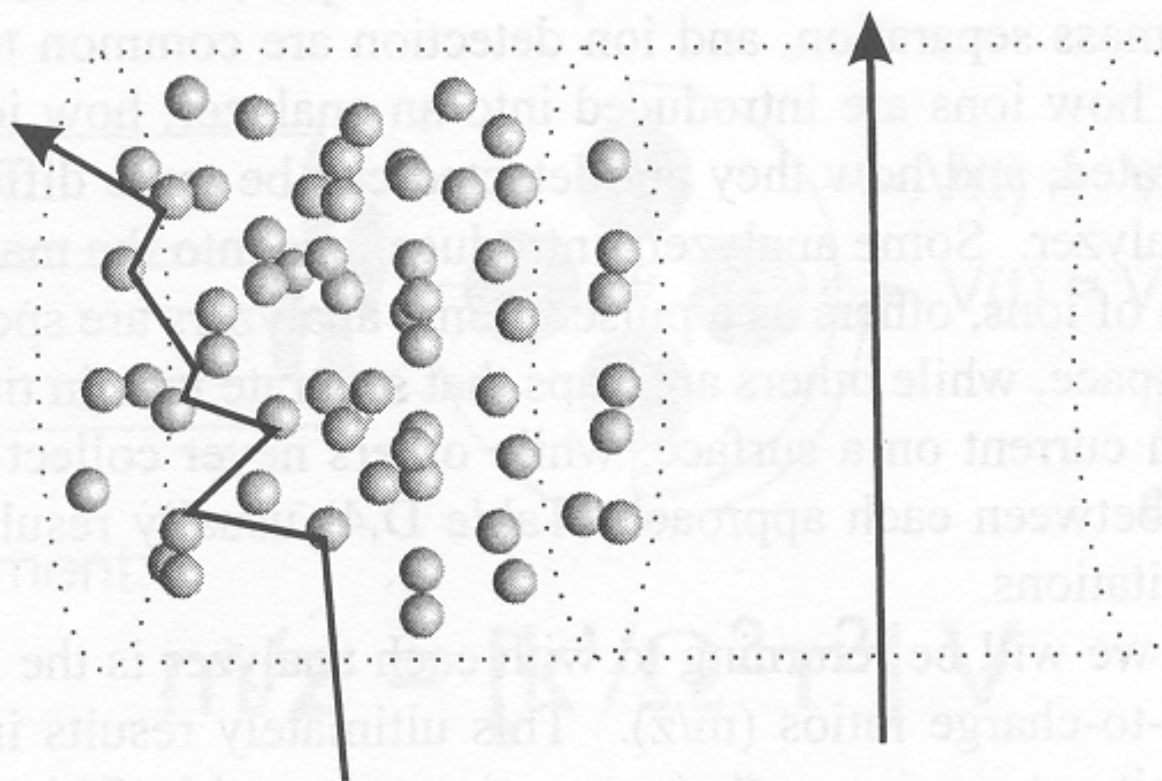
For MS analyzers,  $L \geq 1$  meter then  $p \leq 10^{-5}$  Torr

In practice-  $p \leq 10^{-5}$  Torr

Need very efficient vacuum pumps for MS!

High Pressure

Low Pressure



# Ionization Methods

## Characteristics:

### 1. Energy Imparted:

Soft Ionization (less fragmentation)-

**MALDI- matrix-assisted laser desorption/ionization**

**ESI- electrospray ionization**

Hard- EI (electron impact ionization),  
FAB (fast atom bombardment),  
SIMS (secondary ion mass spectrometry)

### 2. Sample State:

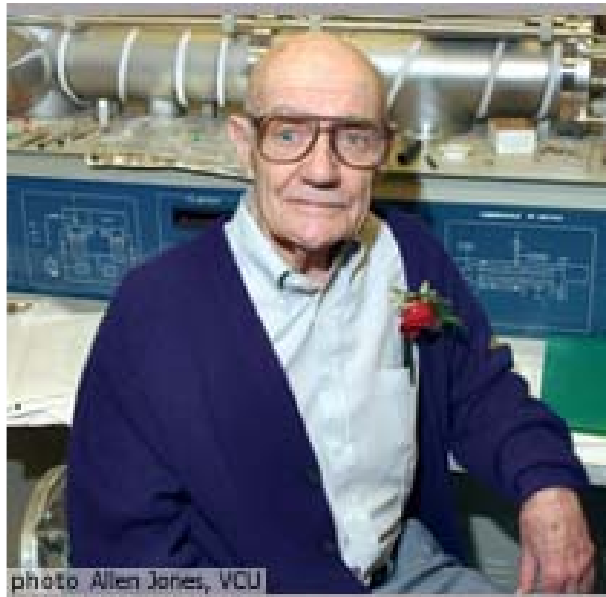
Gas- EI, CI

Liquid- nebulization to introduce droplets, **ESI**, thermospray

Solid- uses an absorbing matrix &  
irradiate with particles or photons  
**MALDI**, FAB, field & plasma desorption

# Ionization Methods: ESI & Laser Desorption

## 2002 Nobel Prize in Chemistry



**John B. Fenn** is the chemist who invented the electrospray method. Today it is used in laboratories all over the world.



**Koichi Tanaka's** idea was to use the energy from laser light, ingeniously transferred to the proteins, to get them to let go of one another and hover freely.

<http://nobelprize.org/chemistry/laureates/2002/illpres/mass.html>



# Ion Source

Recall: Mean Free Path =  $L$  (cm) =  $4.95/p$  for  $p$  in milliTorr

## Problem #1:

Most ionizers Need at  $P = \sim 0.5$  Torr so that  $L = \sim 0.1$  mm

Recall the analyzer is at  $\leq 10^{-5}$  Torr

## Solution #1:

Increase collisions  $\longrightarrow$  Better ionization

Higher pressures  $\longrightarrow$  Increase collisions

ESI- Ionization at atmospheric pressure (760 Torr)  
( $10^3$ - $10^4$  more efficient ionization)

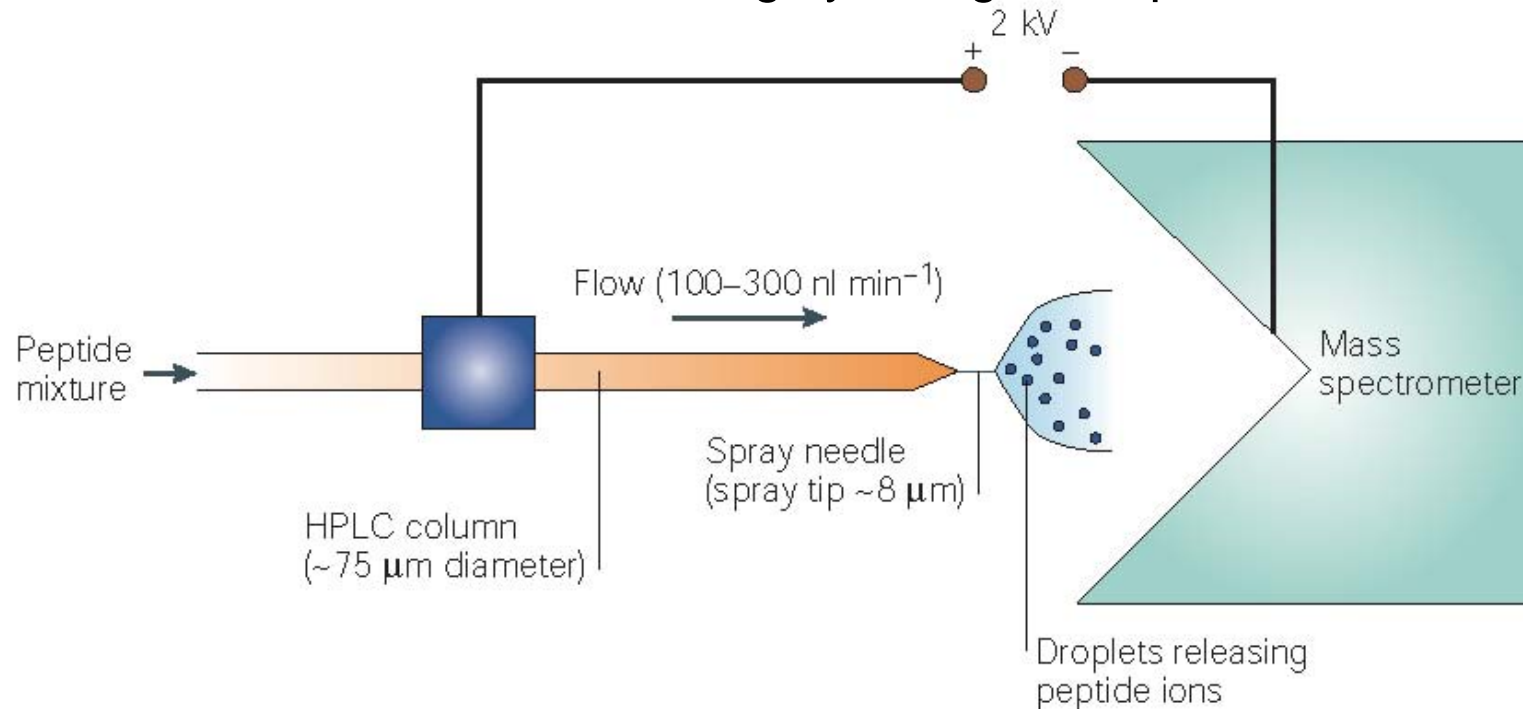
Enabled by:

1. focusing lenses with small openings
2. focusing multipole lenses
3. very high capacity pumps

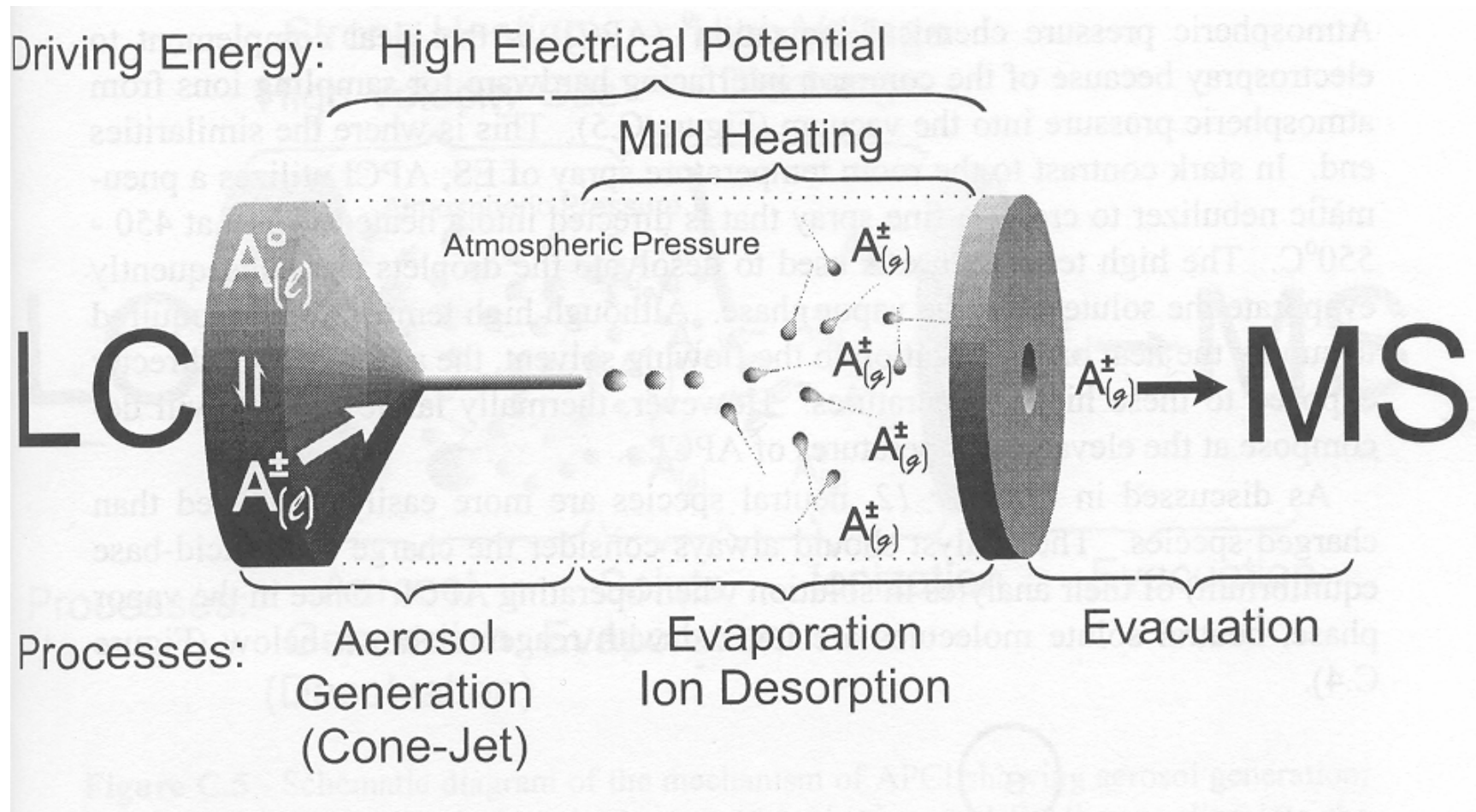
# Electrospray Ionization

1. Capillary with a small orifice or tip.
2. Slowly, flowing liquid at atmospheric pressure.
3. High E field ( $\sim 10^6$  V/m) across the tip.

Charge accumulation at liquid surface of capillary tip.  
Leads to formation of highly charged droplets.



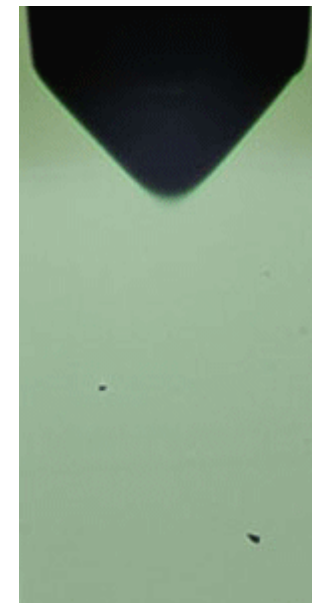
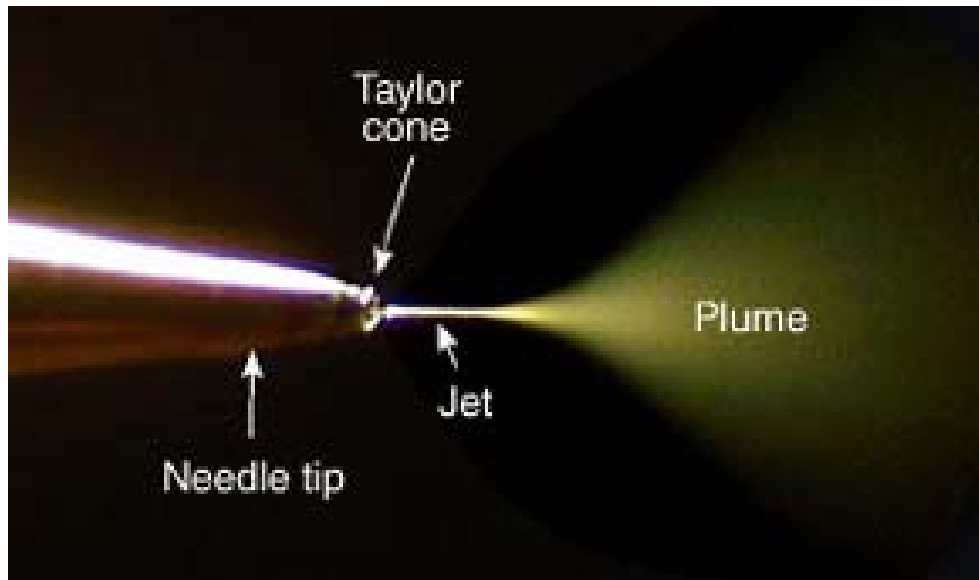
# Taylor Cone & Droplet Formation



# Taylor Cone & Droplet Formation



**Onset Voltage-** Minimum voltage at which droplets form.



<http://www.newobjective.com/electrospray/>

Supplemental Info for  
Marginean et al,  
Anal. Chem. 2004,  
v76, p. 4202-7

# Taylor Cone & Droplet Formation

900 V - no spray

1000 V - Taylor-cone/droplet oscillation, more "drops" than spray

1100 V - cone/droplet oscillation. approx 50% spray

1200 V - cone/droplet oscillation, on the verge of a stable Taylor cone

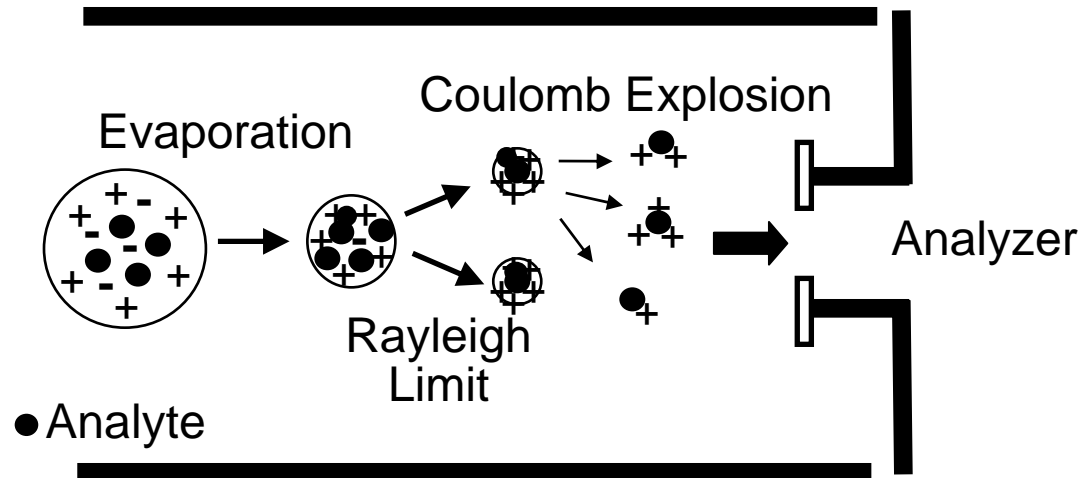
1300 V - stable cone-jet

1400 V - cone-jet on the verge of "jumping", slight instability

1550 V - multiple cone-jets



# Mechanism of Ionization



## Charged Residue Model

1. Solvent evaporates shrinking droplets
2. Coulombic forces exceed cohesive forces
3. Droplet break up

Rayleigh Eqn:  $q^2 = 8\pi^2\epsilon_0\gamma d^3$

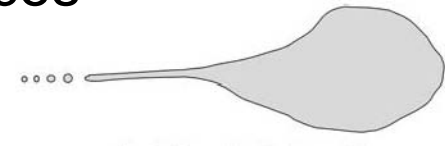
$q$  = charge;

$\gamma$  = surface tension;

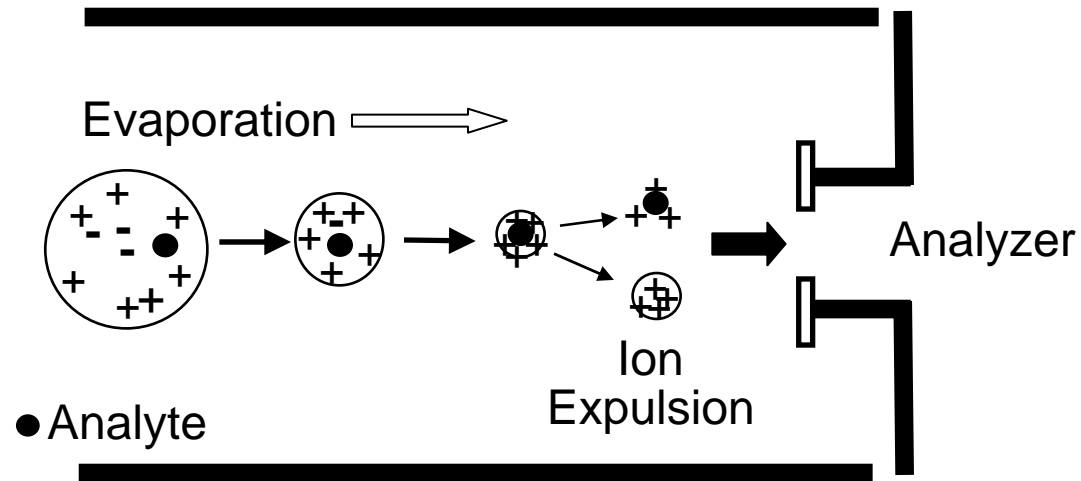
$\epsilon_0$  = environment permittivity

$d$  = droplet diameter

4. Repeats eventually leading to single ions



# Mechanism of Ionization



## Ion Evaporation Model

1. Solvent evaporates shrinking droplets
2. Coulombic forces on ion exceed surface tension
3. Sufficiently large E field at droplet surface results in release of single ions.

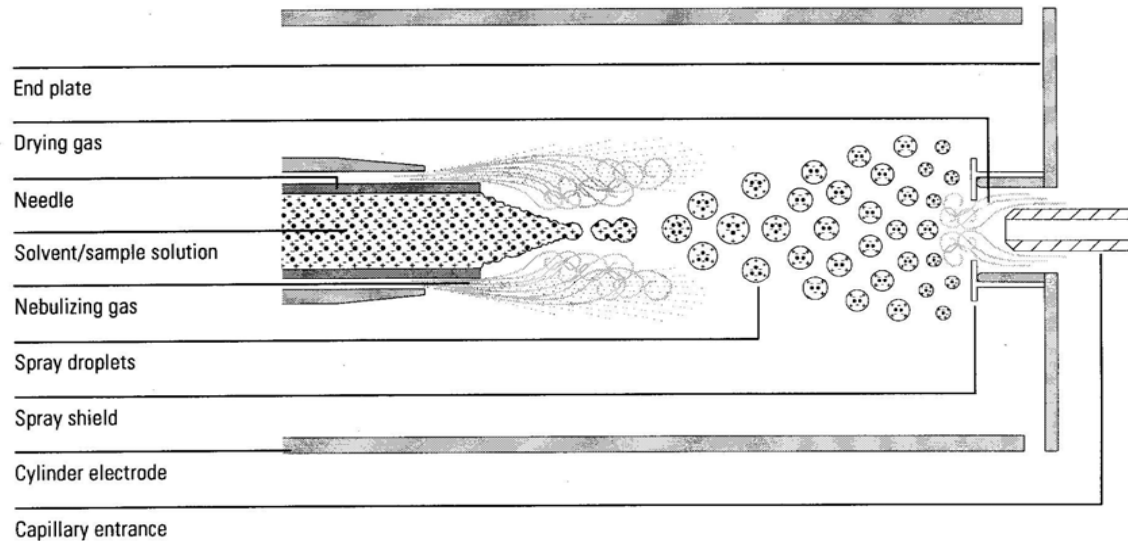
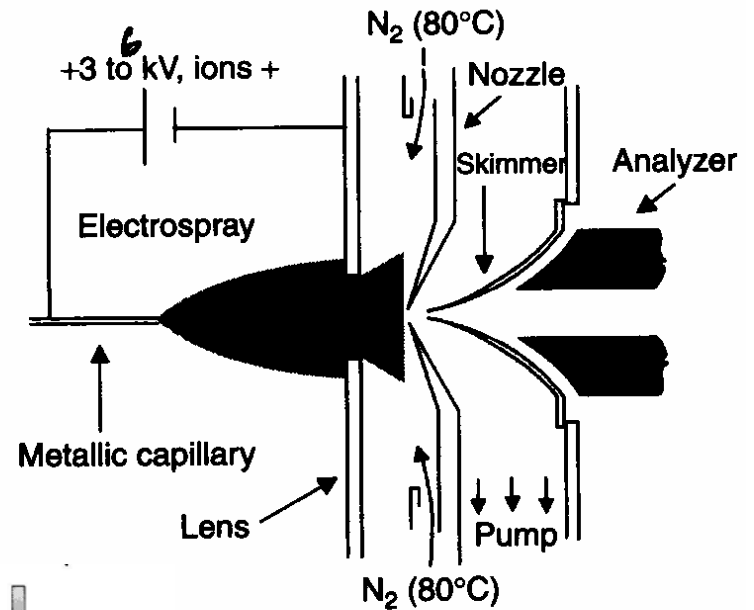
# Electrospray Ionization

## Problem #2:

Sample cooling due to adiabatic expansion can yield ion clusters of sample.

**Solution #2:** Evaporate residual solvent

**Employ:** 1. Hi temp. transfer tube or,  
2. Heated  $N_2$  counter flow or,  
3. Heated gas curtain





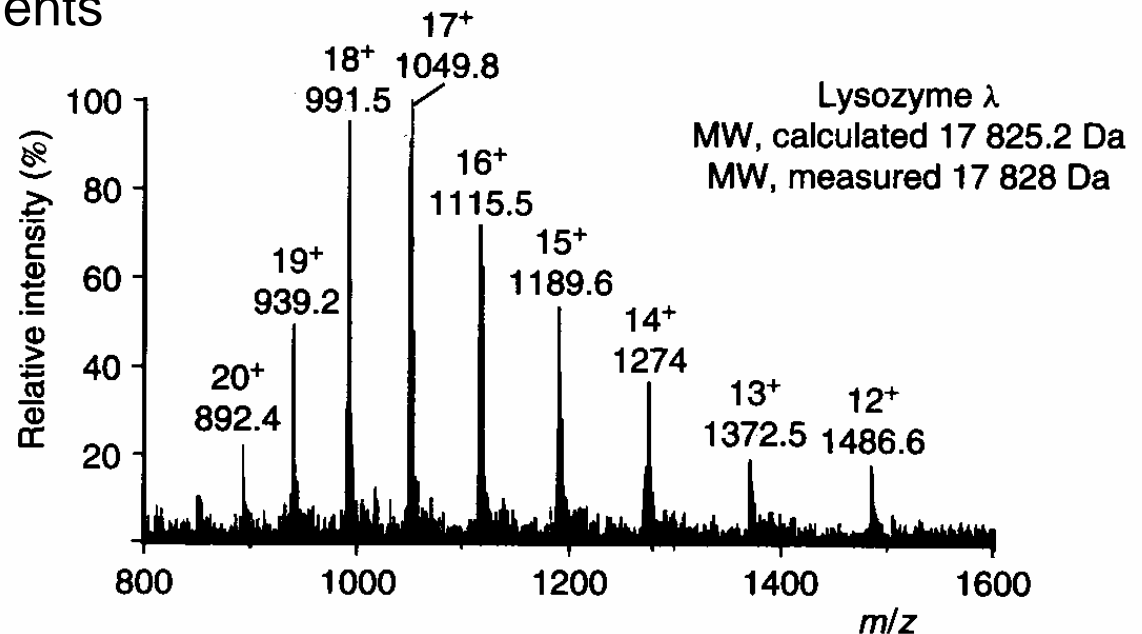
# Analyte Charge

1. Ion formation due to:
  - a. charge on droplets
  - b. electrochemical processes
  - c. adduct formation ( $\text{Na}^+$ ,  $\text{NH}_4^+$ )
  - d. gas-phase reactions
2. Produce multiply charged ions from large molecules  
Proteins-  $\sim 1$  charge/1 kD on average

3. Multiple mass measurements  
due to multiple  $m/z$

protonation-  $(M + z\text{H})^{z+}$   
deprotonation-  $(M - z\text{H})^{z-}$

4. Very little fragmentation



# Charge Distribution

**For a peak located at  $m_1$  i.e. peak 1:**

$$z_1 m_1 = M + z_1 m_p \quad \text{for} \quad \begin{array}{l} M = \text{mass of a molecular ion} \\ m_1 = \text{measured mass-to-charge} \\ z_1 = \# \text{ of added protons on } m_1 \text{ peak} \\ m_p = \text{mass of proton} \end{array}$$

**Consider a peak separated from peak 1 by  $j-1$  peaks:**

$$(z_1 - j) m_2 = M + (z_1 - j) m_p \quad \text{for} \quad \begin{array}{l} m_2 = \text{measured mass-to-charge} \\ (z_1 - j) = \# \text{ of protons on the peak at } m_2 \end{array}$$

**Then for peak 1:**

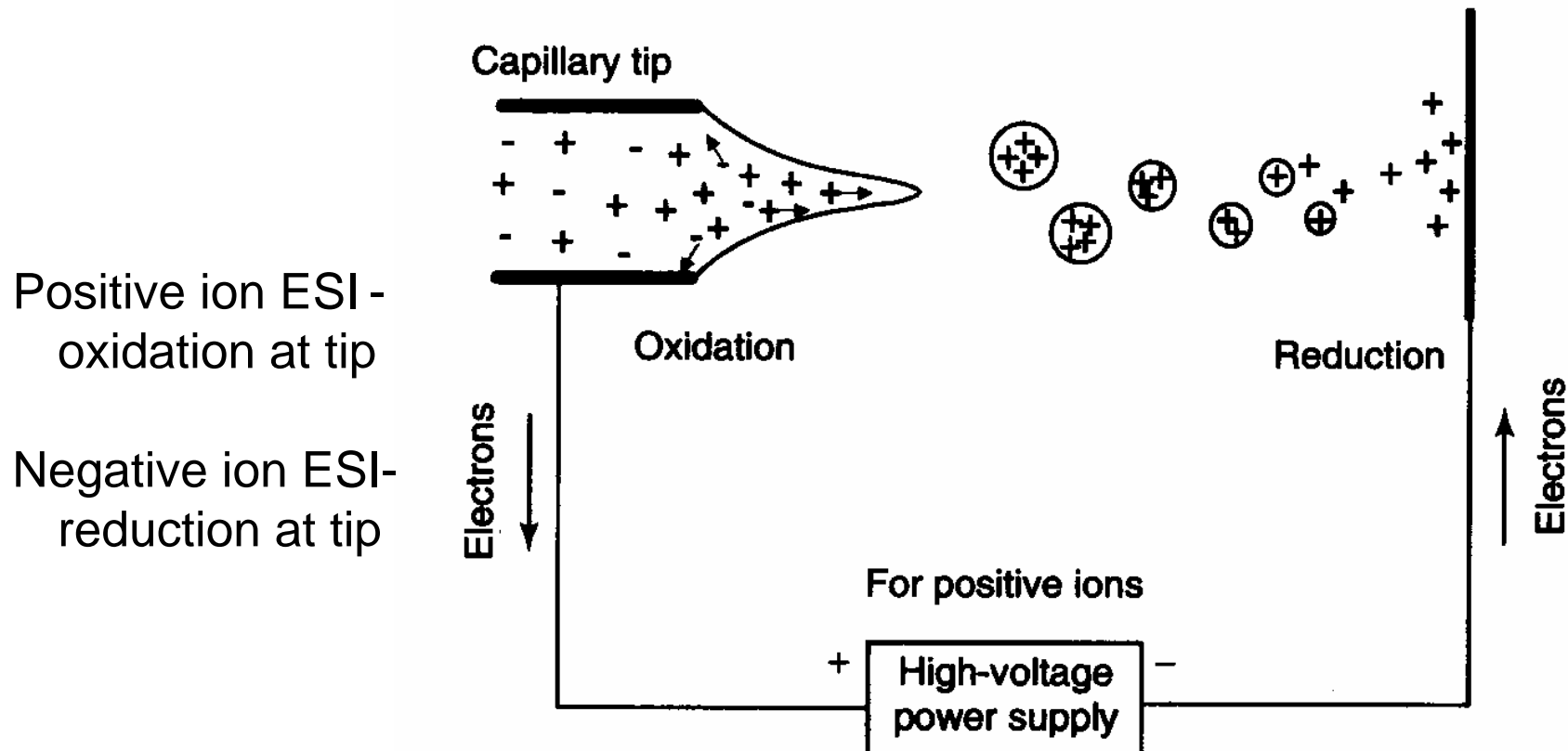
$$z_1 = j (m_2 - m_p) / (m_2 - m_1) \quad \& \quad M = z_1 (m_1 - m_p)$$

**Multiply charged species permit calculation of  $z$  &  $M$  in ESI!**

# Electrochemistry in ESI

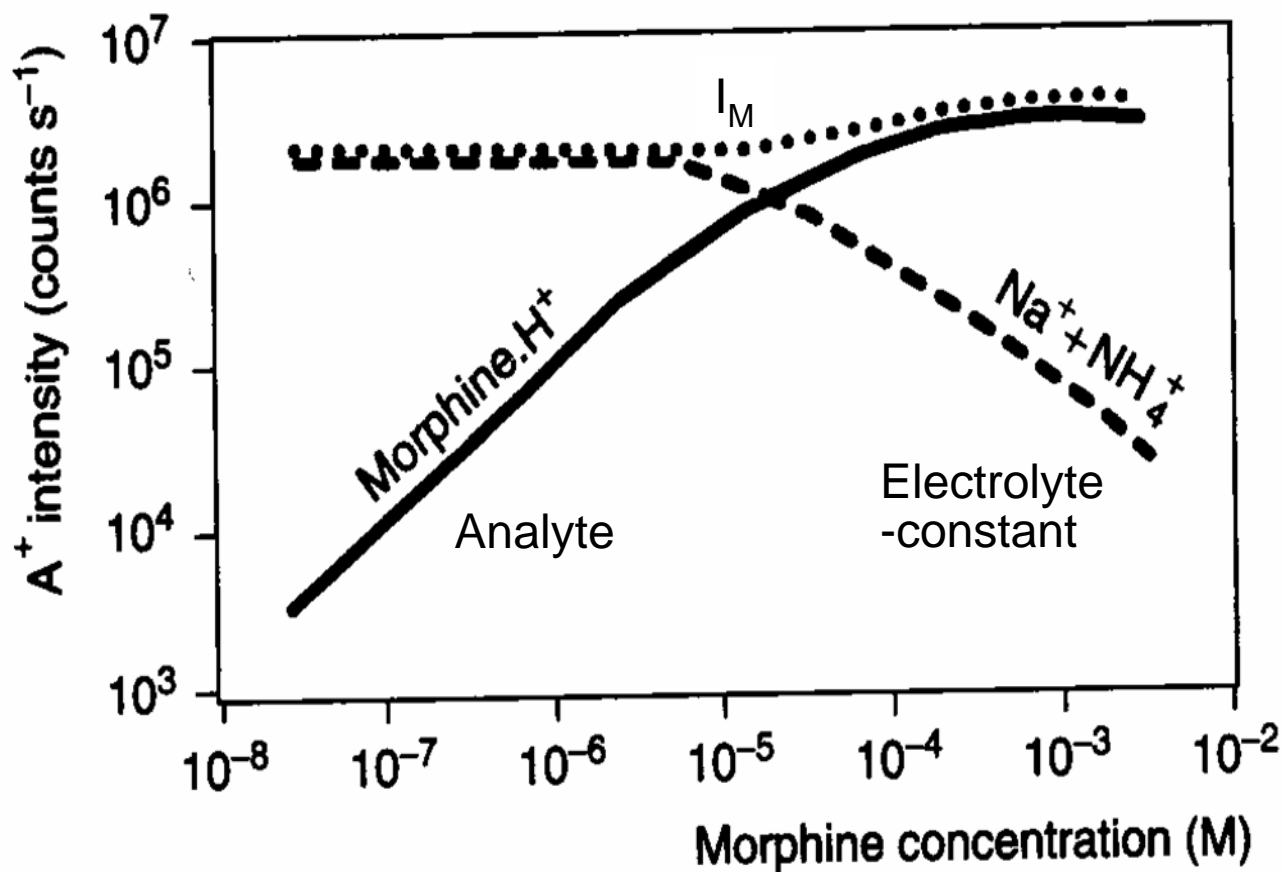
Recall: Ion charge is due to electrochemical processes at probe the tip

**Total # of ions/time into the spectrometer is limited by the current at the probe tip.**



# Electrochemistry in ESI

$I_M$  = Limiting Ion Current (total current is limited by oxidation)  
(typically  $\sim 1 \mu\text{A}$ )

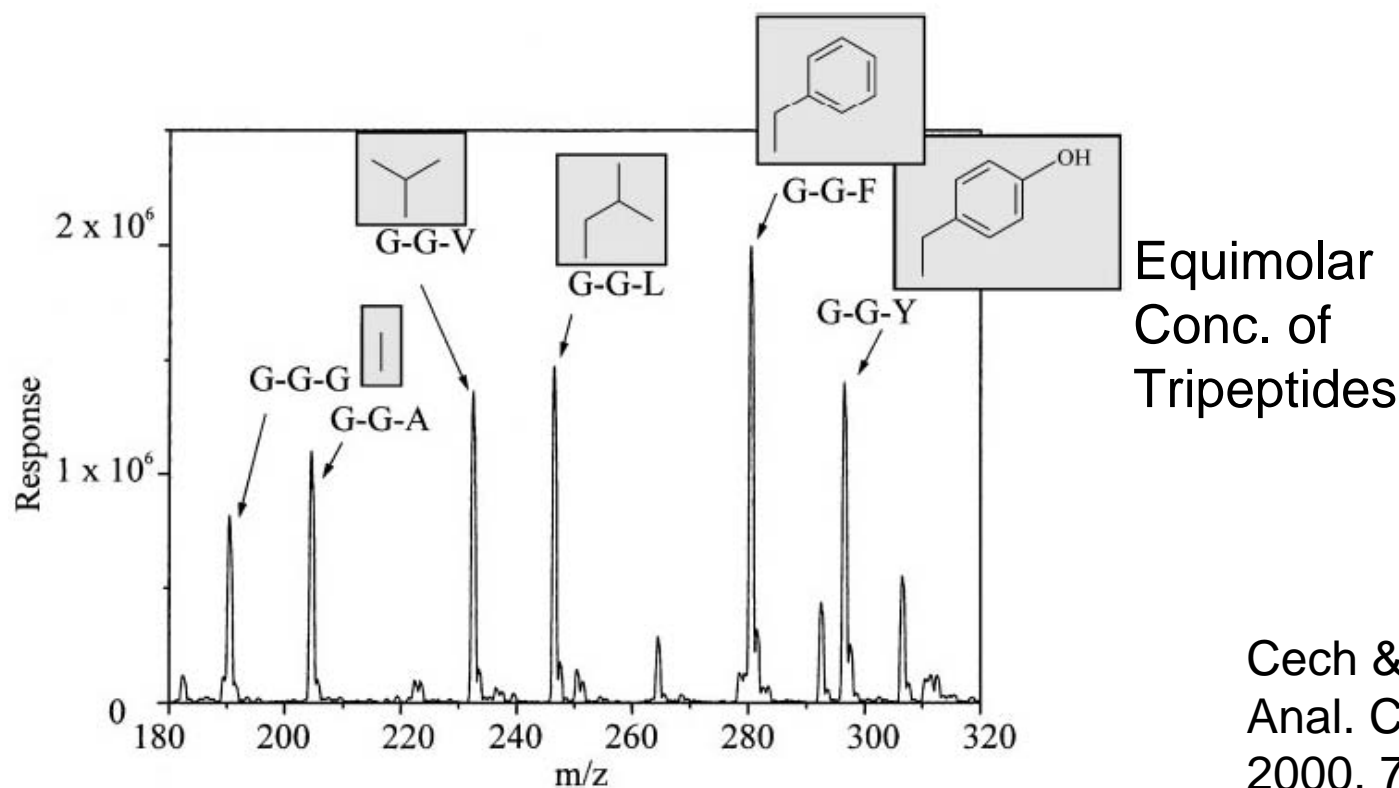


# Surface Activity in ESI

**Equilibrium Partitioning Model:** Molecules that prefer the surface of the droplets have a better ESI response.

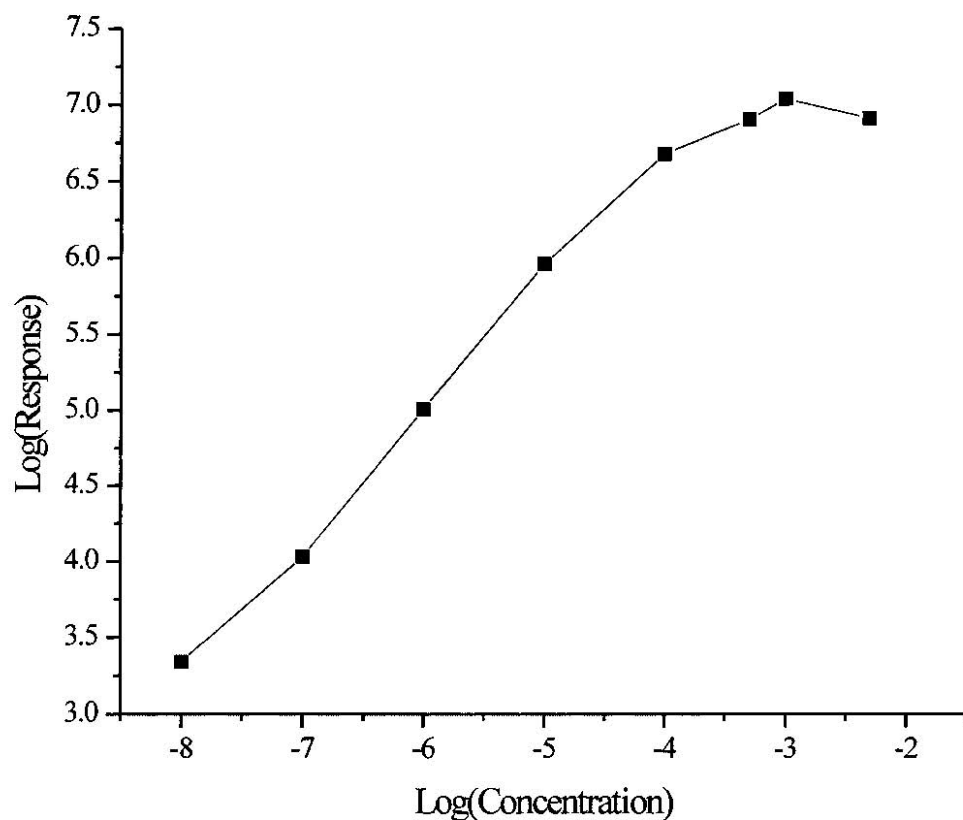
Partitioning Coefficient ( $k$ ) =  $[\text{Analyte}]_{\text{Surface}}/[\text{Analyte}]_{\text{Interior}}$

$k$  depends on polarity, charge density, basicity.



Cech & Enke,  
Anal. Chem.  
2000, 72:2717.

# Analyte Concentration for ESI



Cech & Enke,  
Mass Spec. Rev.  
2001, 20:362-87.

Typically:  $10^{-8} \text{ M} < [\text{Analyte}] < 10^{-5} \text{ M}$

# Microspray and Nanospray

**Table 3.2. Experimental conditions that distinguish electrospray, microspray, and nanospray ionization**

	Electrospray <sup>1</sup>	Microspray <sup>1</sup>	Nanospray <sup>2</sup>
<i>Typical flow rate</i>	2 $\mu\text{L}/\text{min}$	0.2 $\mu\text{L}/\text{min}$	0.02 $\mu\text{L}/\text{min}$
<i>Needle size (i.d.)</i>	75 $\mu\text{L}/\text{min}$	75 $\mu\text{m}$	5 $\mu\text{m}$
<i>Approximate limits of detection</i>	10 femtomole	1 femtomole	50 attomole
<i>Sheath liquid flow required</i>	Yes	No	No

# ESI

## Points:

1. Simple
2. Good for large, nonvolatile biomolecules  
(up to 50 kD)
  - a.  $10 \text{ nM} < \text{Analyte} < 10 \text{ } \mu\text{M}$
  - b. Liquid solutions
  - c. Must desalt sample
3. Easily coupled to HPLC,  $\mu$ -HPLC, and CE
  - a. Flow rates of  $\mu\text{l/min}$ - $\text{nl/min}$
  - b. Also coupled to microfluidic chips
4. Little fragmentation
  - a. Easy to determine M.W.



# MALDI- Karas & Hillenkamp

## 1. Sample + Matrix

Matrix- small organic molecule absorbing at laser  $\lambda$ .

## 2. Dried- Matrix crystals surround the analyte.

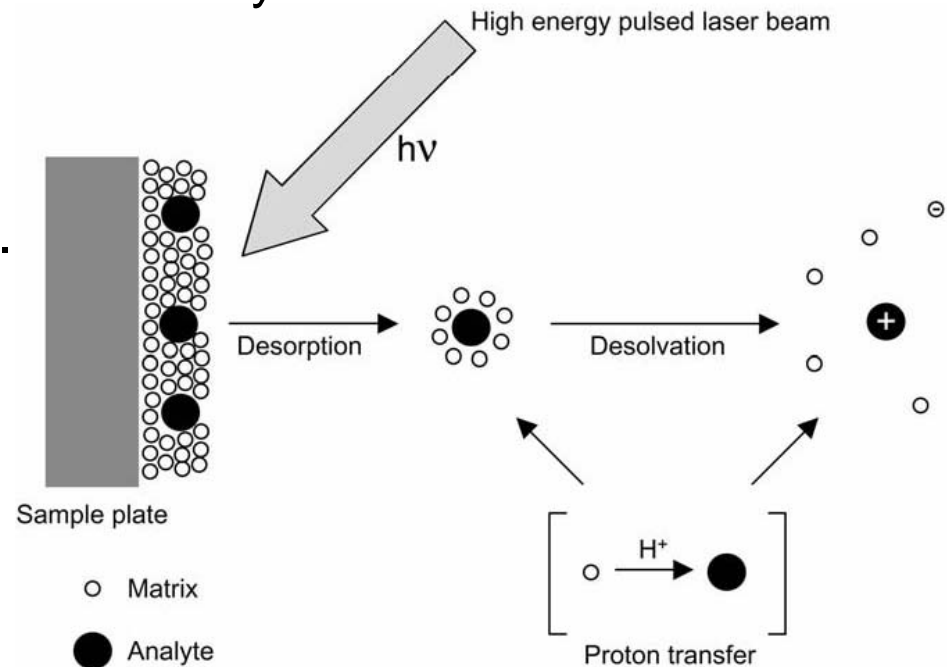
## 3. Ablate with pulsed laser

a. Rapid matrix heating with localized sublimation.

b. Gas phase matrix brings analyte along.

c. Very little energy transferred to analyte.

d. Ionization occurs by gas-phase proton transfer by photoionized matrix.

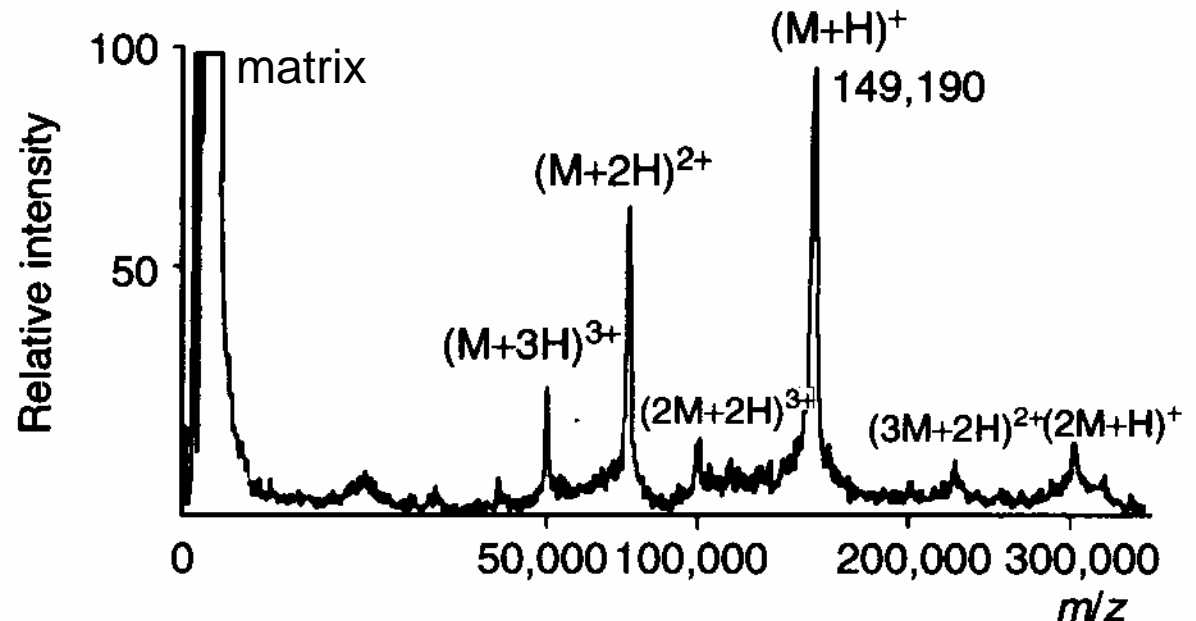


Lane,  
Cell Molec. Life Sci.  
2005, 62:848-869.

# Role of the Matrix

1. Minimizes sample clustering  
matrix molecules  $\gg$  analyte molecules
2. Absorbs laser energy minimizing sample damage  
Mainly the monocharged species i.e.  $(M+H)^+$ ,  $(M+Na)^+$ , etc
3. Increases efficiency of energy transfer  
Enhances sensitivity- picomoles of peptides & proteins up to 300 kD
4. Improves analyte independence

MALDI MS of  
antibody



Hillenkamp & Karas Meth.  
Enzym. 1990, 193: 280-295.

# Fragmentation in MALDI

**Threshold Irradiance-** Minimum laser power resulting in matrix desorption.  
At > threshold- increased fragmentation, decreased resolution

**Lasers-** UV ( $N_2$ , 337 nm; Nd:YAG, 260 or 355 nm)  
IR (Er:YAG, 2.94  $\mu m$ ;  $CO_2$ , 10.6  $\mu m$ )

## Fragmentation:

1. Prompt- On sample surface during desorption
2. Fast- At source but after desorption & before acceleration
3. Post Source Decay- Occurs after onset of acceleration

Seen in MALDI  
Spectra

Peak broadening with decreased  
resolution & sensitivity.

# MALDI Tolerates Impurities

**Table 3.3. Tolerance limits of matrix-assisted laser desorption/ionization for various reagents used in protein and peptide isolation protocols.**

Reagent	Approximate maximum tolerable concentration
Urea	0.5 M
Guanidine	0.5 M
Sodium dodecyl sulfate (SDS)	0.01 %
Detergents other than SDS	0.1 %
Dithiothreitol	0.5 M
Glycerol	1 %
Alkali metal cations such as Na <sup>2+</sup>	0.5 M
Tris	50 mM
Phosphate	50 mM

# MALDI Summary

## Points:

1. Best mass accuracy- 0.01% (1 Da in 10 kDa)
2. Very fast and inexpensive (MALDI-TOF)
3. Can do large molecules ~300,000 kDa
4. Little fragmentation  
Good for determining molecular weights
5. Very sensitive
6. Nontrivial to couple to separation techniques
7. Frequently used with TOF analyzer  
for "Peptide Mass Fingerprinting"

# References

1. Caprioli, R, Sutter, M, Stoeckli, M. Online Mass Spectrometry Tutorial  
<http://staging.mc.vanderbilt.edu/msrc/tutorials/ms/ms.htm>
2. Standing KG. 2003. Peptide and protein de novo sequencing by mass spectrometry. *Current Opinion in Structural Biology*. 13:595-601.
3. Lane CS. 2005. Mass spectrometry-based proteomics in the life sciences. *Cellular and Molecular Life Sciences*. 62:848-869.
4. Stutz H. 2005. Advances in the analysis of proteins and peptides by capillary electrophoresis with matrix-assisted laser desorption/ionization and electrospray-mass spectrometry detection. *Electrophoresis*. 26:1254-1290.
5. Cech NB, Enke CG. 2001. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews*. 20:362-387.
6. Stump MJ, Fleming RC, Gong WH, Jaber AJ, Jones JJ, Surber CW, Wilkins CL. 2002. Matrix-assisted laser desorption mass spectrometry. *Applied Spectroscopy Reviews*. 37:275-303.
7. Steen H, Mann M. 2004. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews: Molecular Cell Biology*. 5:699-711.
8. Herbert CG, Johnstone RAW. *Mass Spectrometry Basics*. 2003. CRC Press, Boca Raton.
9. de Hoffman, E. Stroobant, V. *Mass Spectrometry Principles and Application*. 2001. 2nd Ed., John Wiley & Sons, New York.