

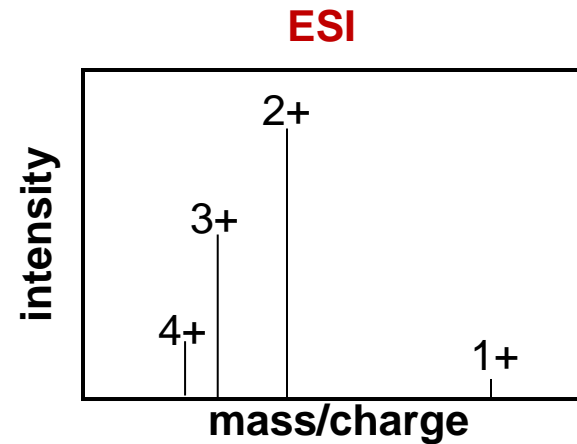
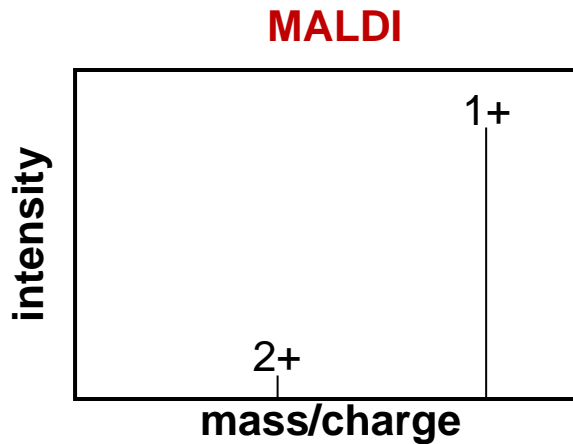
## Proteomics Informatics -

**Analysis of mass spectra: signal processing, peak finding, and isotope clusters (Week 3)**

---

# Charge-State Distributions

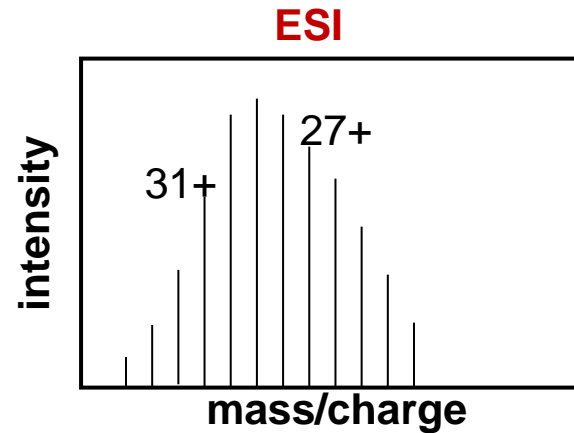
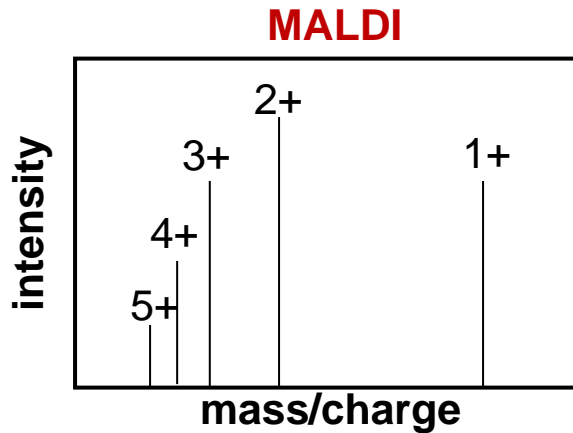
Peptide



$$\frac{m}{z} = \frac{M + nH}{n}$$

M - molecular mass  
n - number of charges  
H - mass of a proton

Protein



# Charge-State

---

$$\frac{m}{z} = \frac{M + nH}{n}$$

M - molecular mass  
n - number of charges  
H - mass of a proton

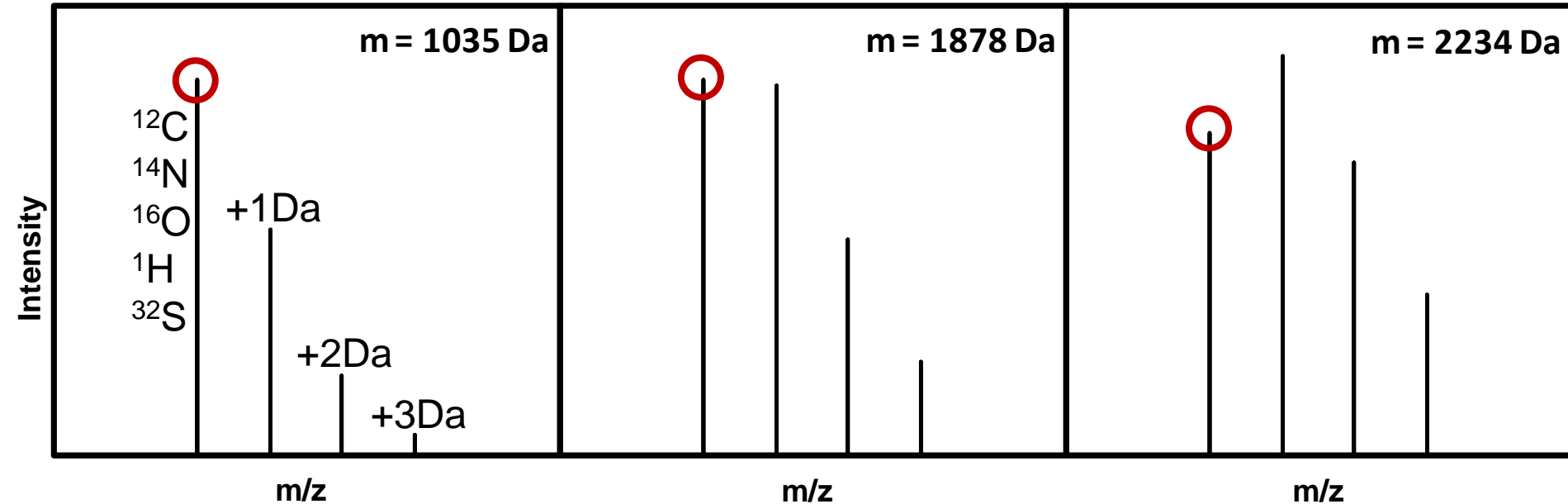
Example:

peptide of mass 898 carrying 1  $H^+$  =  $(898 + 1) / 1 = 899 \text{ m/z}$

carrying 2  $H^+$  =  $(898 + 2) / 2 = 450 \text{ m/z}$

carrying 3  $H^+$  =  $(898 + 3) / 3 = 300.3 \text{ m/z}$

# Isotope Distributions



0.015%  $^2\text{H}$   
1.11%  $^{13}\text{C}$   
0.366%  $^{15}\text{N}$   
0.038%  $^{17}\text{O}$ , 0.200%  $^{18}\text{O}$ ,  
0.75%  $^{33}\text{S}$ , 4.21%  $^{34}\text{S}$ , 0.02%  $^{36}\text{S}$

**Only  $^{12}\text{C}$  and  $^{13}\text{C}$ :**

$p=0.0111$

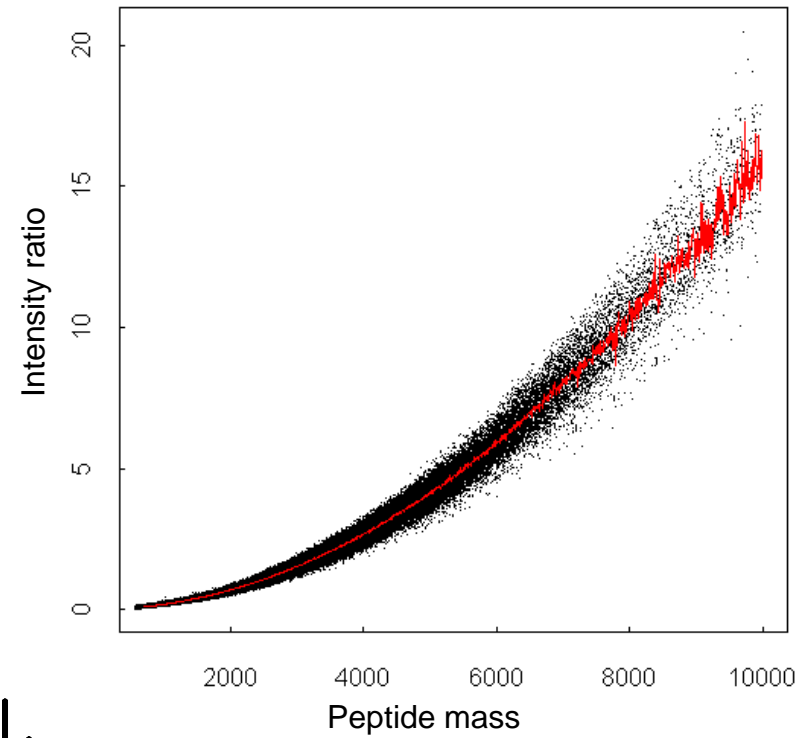
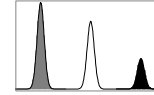
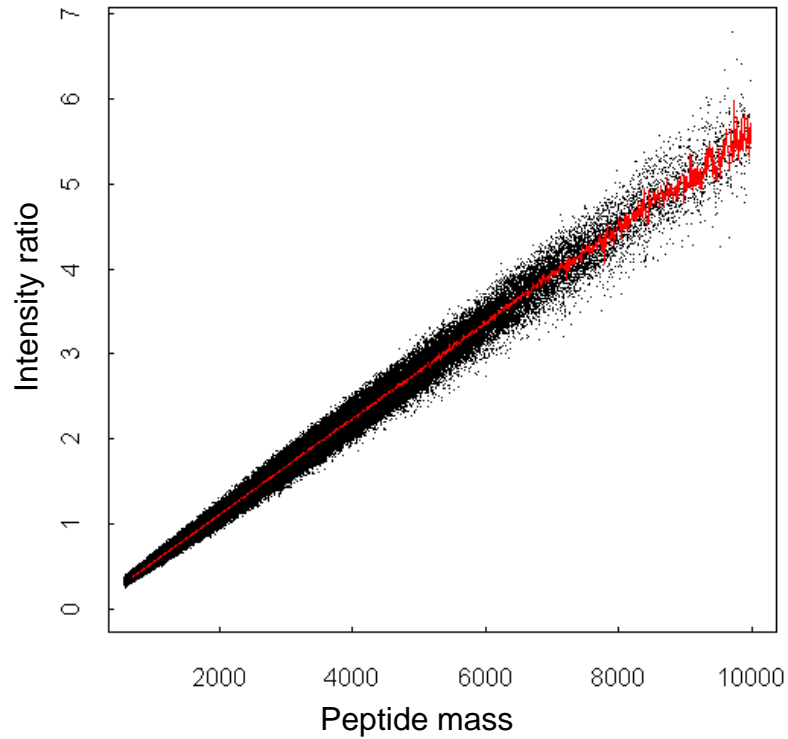
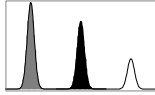
$n$  is the number of C in the peptide

$m$  is the number of  $^{13}\text{C}$  in the peptide

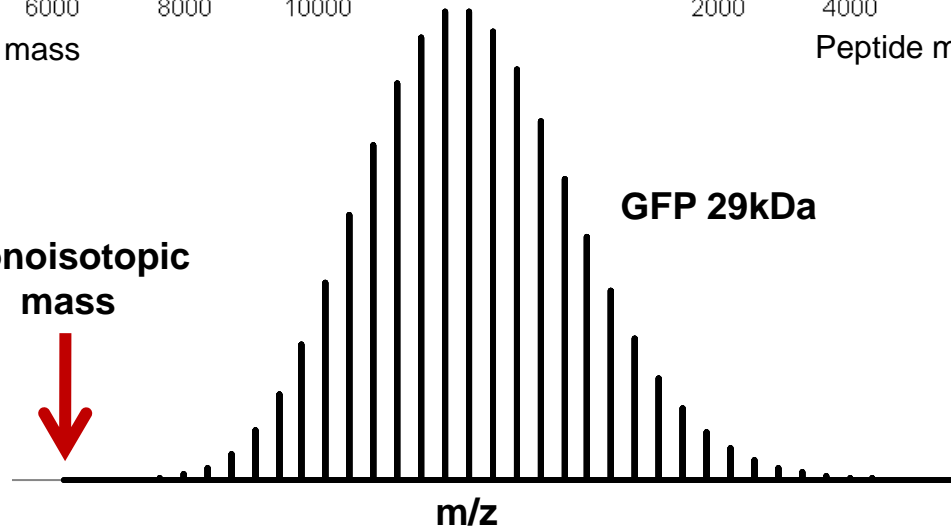
$T_m$  is the relative intensity of  
the peptide  $m$   $^{13}\text{C}$

$$T_m = \binom{n}{m} p^m (1 - p)^{n-m}$$

# Isotope distributions



monoisotopic  
mass



GFP 29kDa

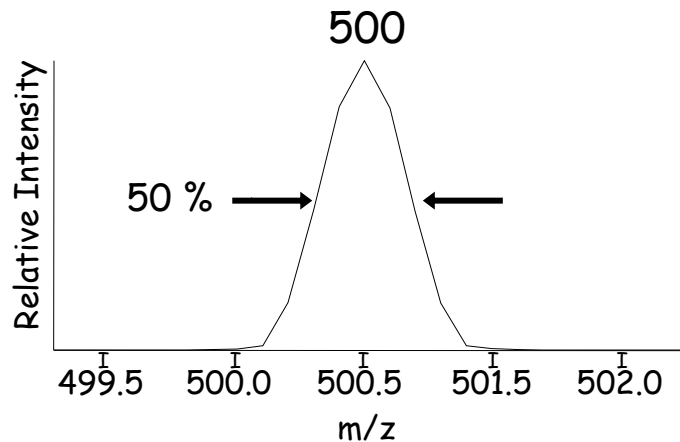
# Resolution

---

$$R = \frac{M}{\Delta M} = \text{resolving power}$$

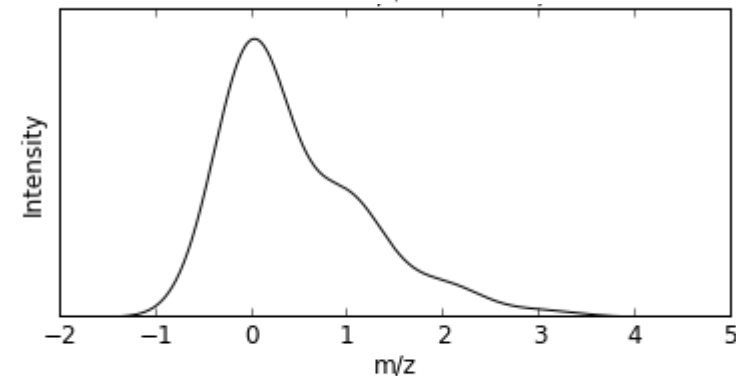
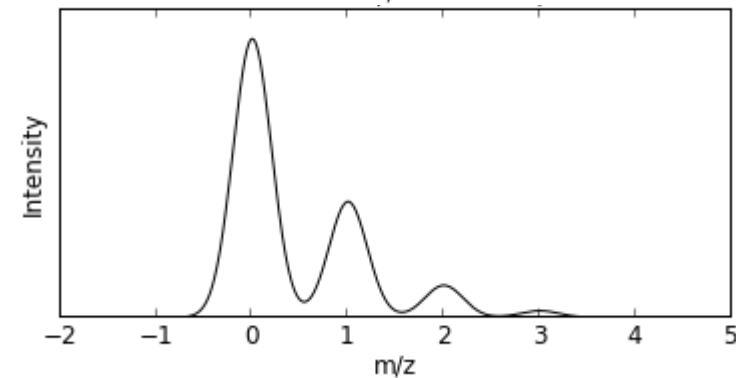
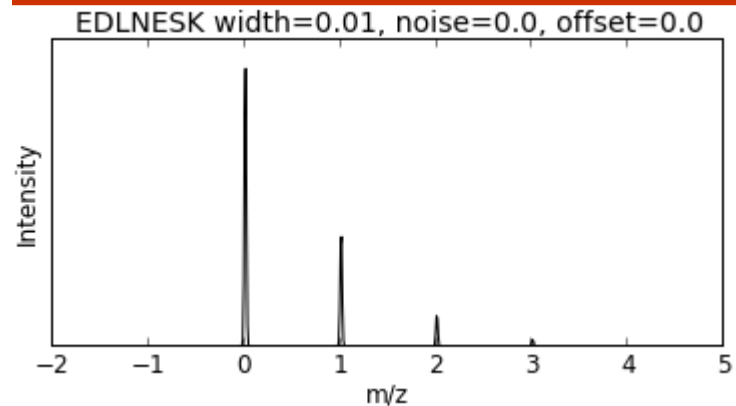
Resolution = minimum peak separation,  $\Delta M$ ,  
which allows to distinguish two ion species

$\Delta M$  = full width at half maximum  
(FWHM)

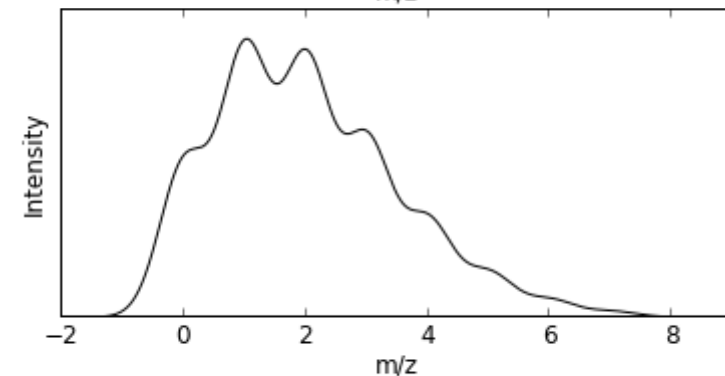
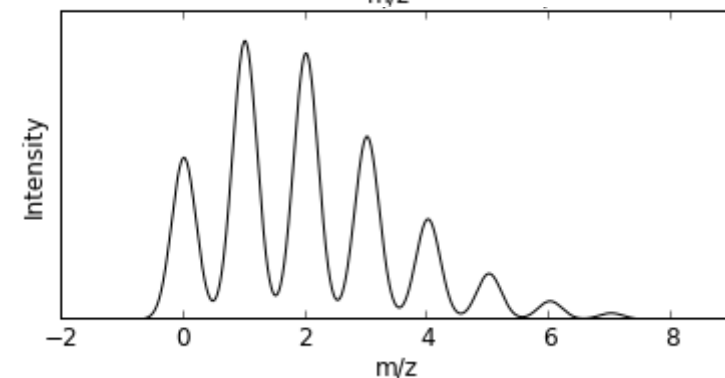
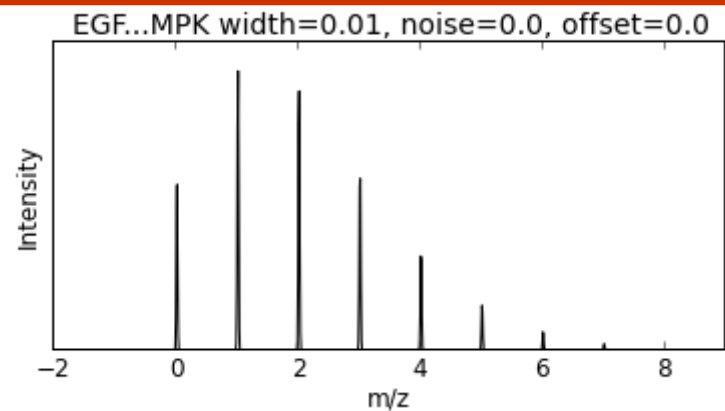


$$\text{Resolution} = M/\Delta M = 500/0.5 = 1000$$

# Resolution



peptide mass = 833.37668  
m/z = 834.38395 for z=1



peptide mass = 3003.3572  
m/z = 3004.3644 for z=1

# Resolution

---

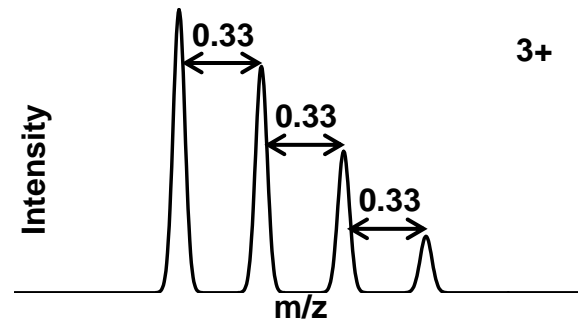
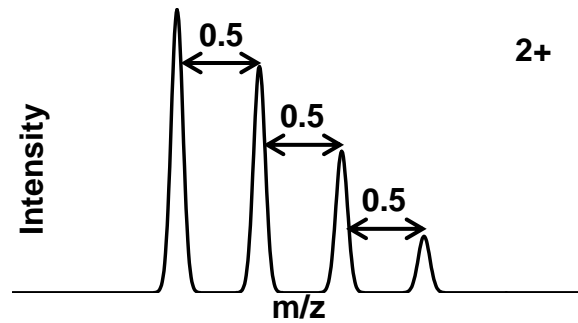
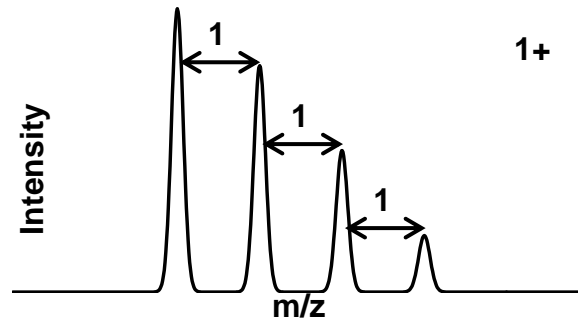
$$R = \frac{M}{\Delta M} = \text{resolving power}$$

- What resolution do we need to differentiate a 1600 Da peptide that carries either an acetylation (+ 42.0100) or trimethylation (42.0464 )?
- $R = 1600/0.0364 = 43,956$



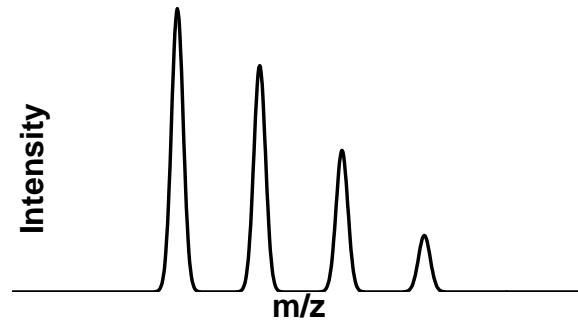
# Isotope Clusters and Charge State

---



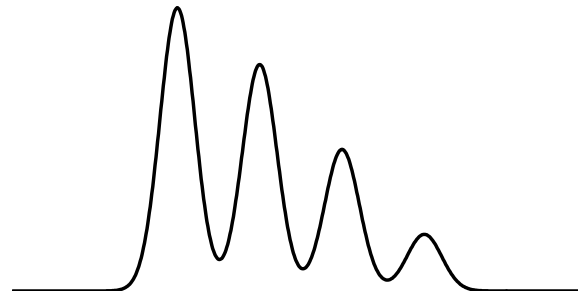
# Isotope Clusters and Charge State

---

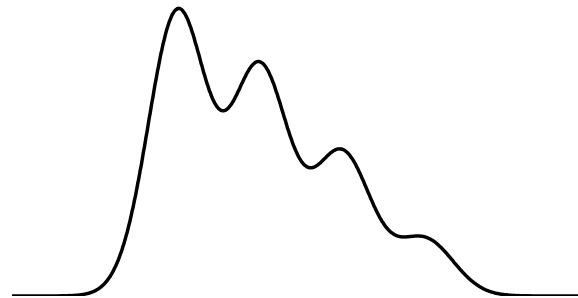


Possible to  
Determine Charge?

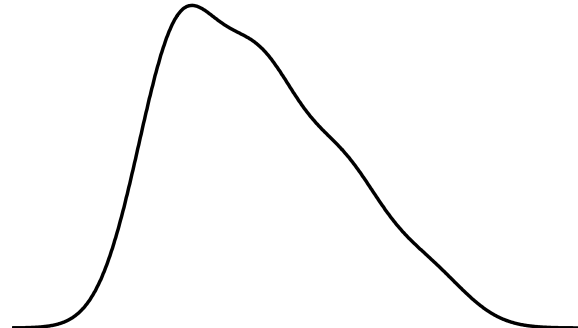
Yes



Yes



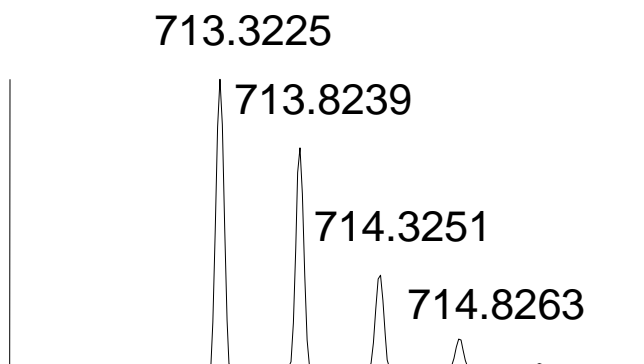
Maybe



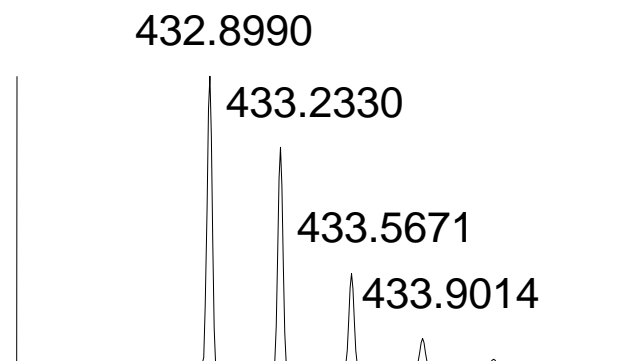
No

# What is the Charge State?

---



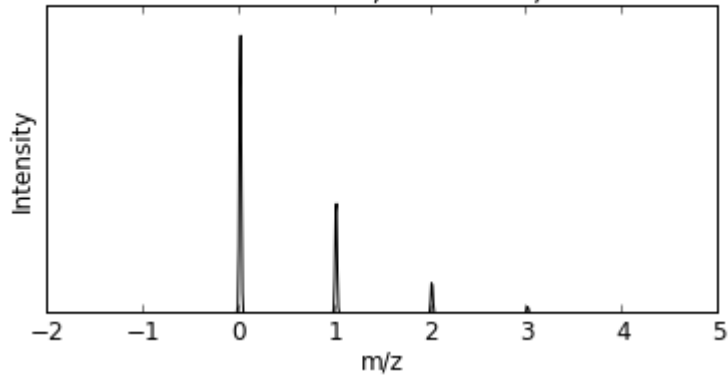
$\Delta$  between the  
isotopes is 0.5 Da



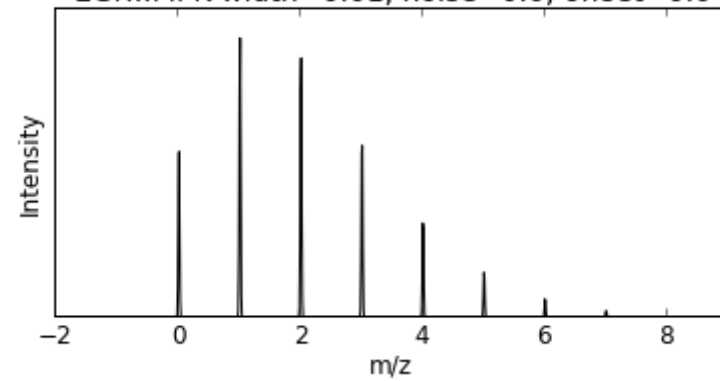
$\Delta$  between the  
isotopes is 0.33 Da

# Noise

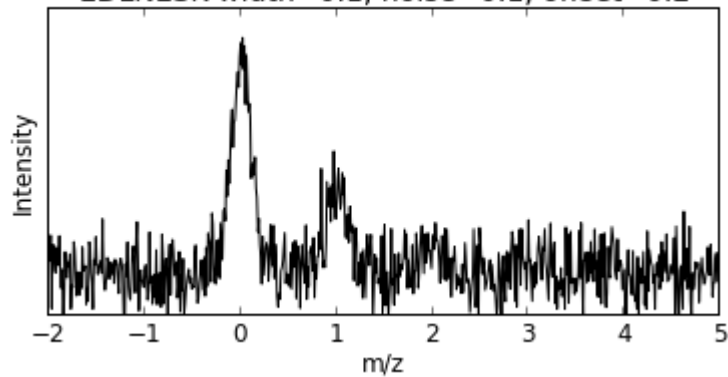
EDLNESK width=0.01, noise=0.0, offset=0.0



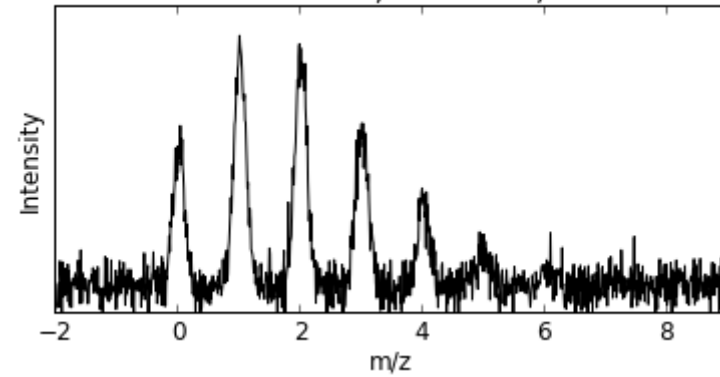
EGF...MPK width=0.01, noise=0.0, offset=0.0



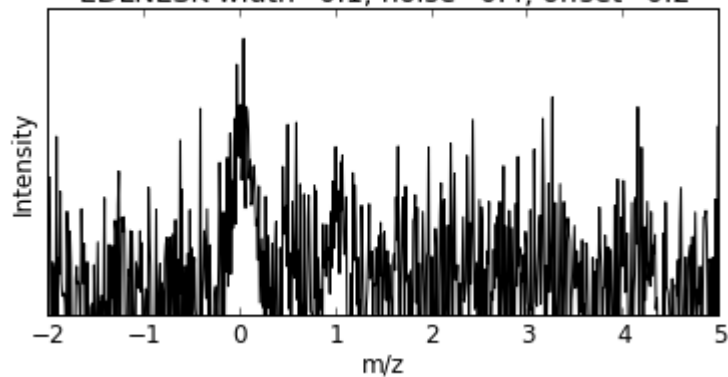
EDLNESK width=0.1, noise=0.1, offset=0.2



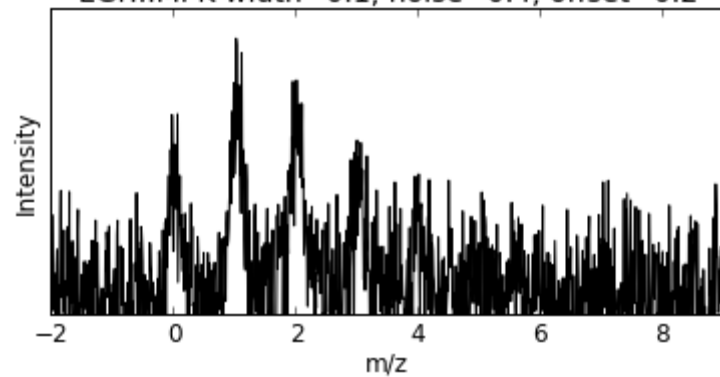
EGF...MPK width=0.1, noise=0.1, offset=0.2



EDLNESK width=0.1, noise=0.4, offset=0.2

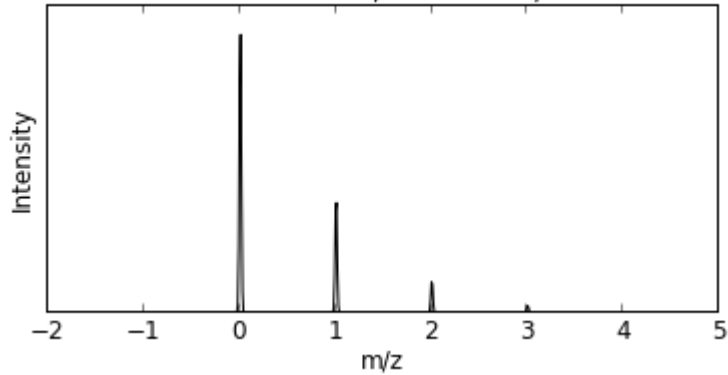


EGF...MPK width=0.1, noise=0.4, offset=0.2

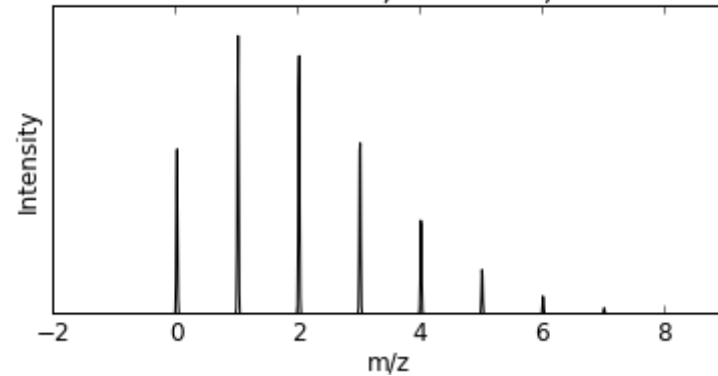


# Smoothing

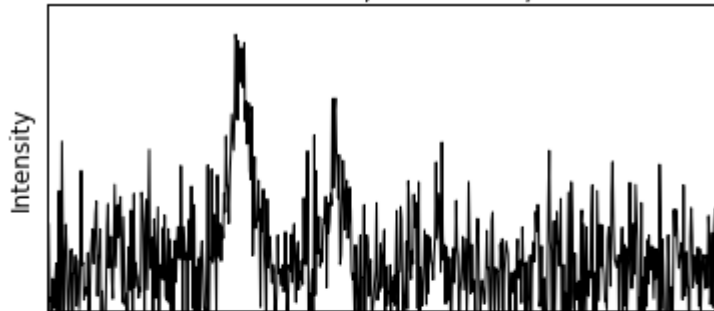
EDLNESK width=0.01, noise=0.0, offset=0.0



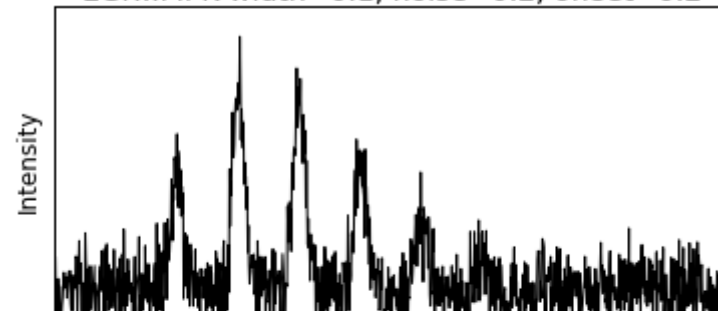
EGF...MPK width=0.01, noise=0.0, offset=0.0



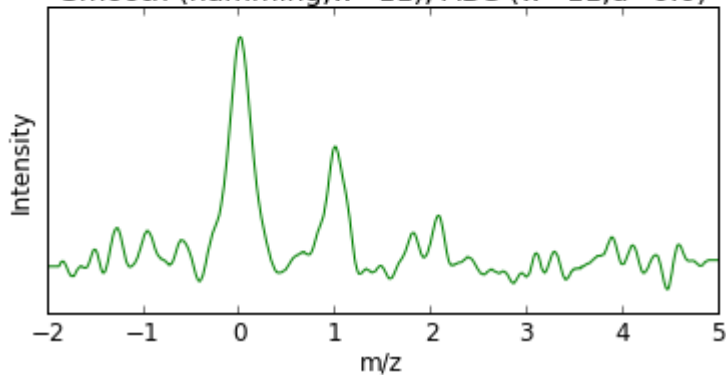
EDLNESK width=0.1, noise=0.2, offset=0.2



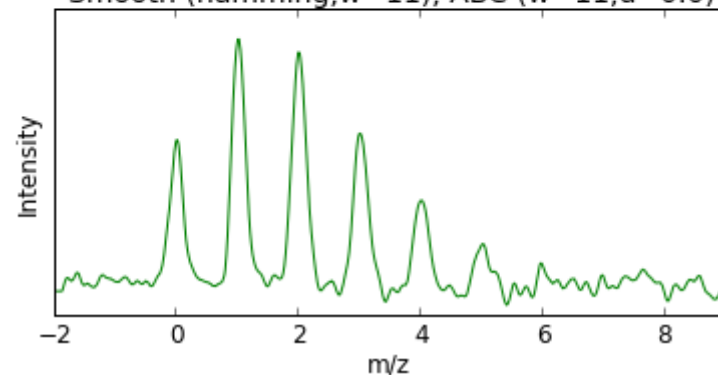
EGF...MPK width=0.1, noise=0.2, offset=0.2



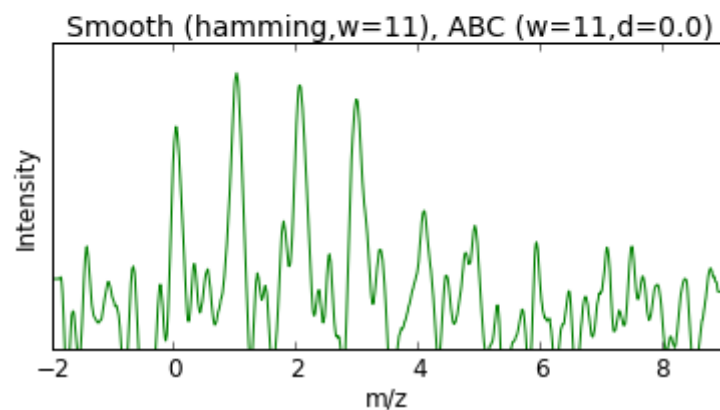
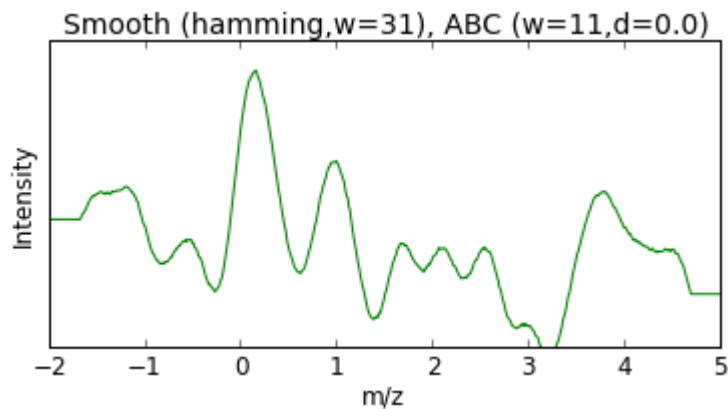
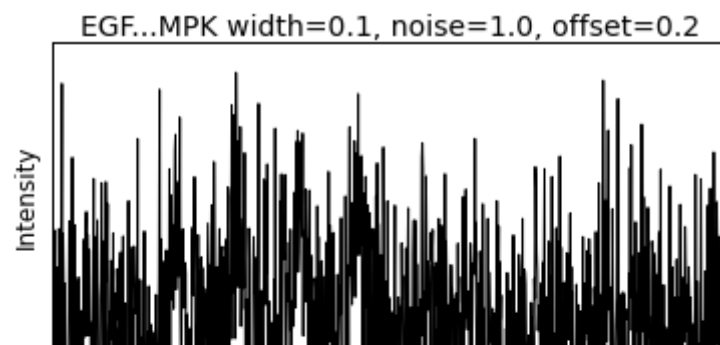
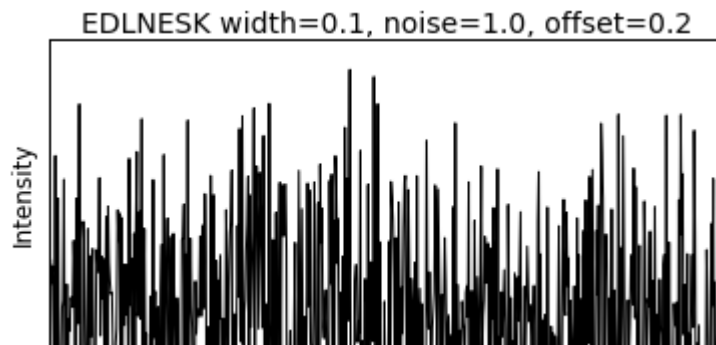
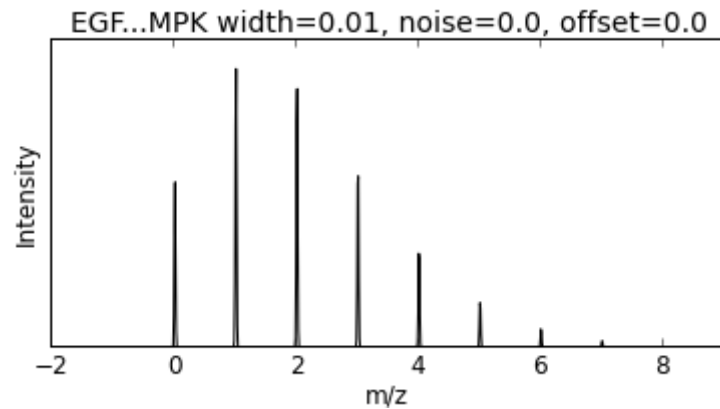
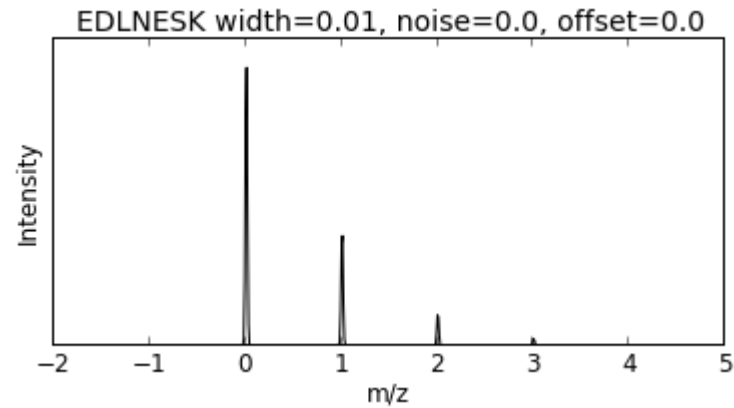
Smooth (hamming,w=11), ABC (w=11,d=0.0)



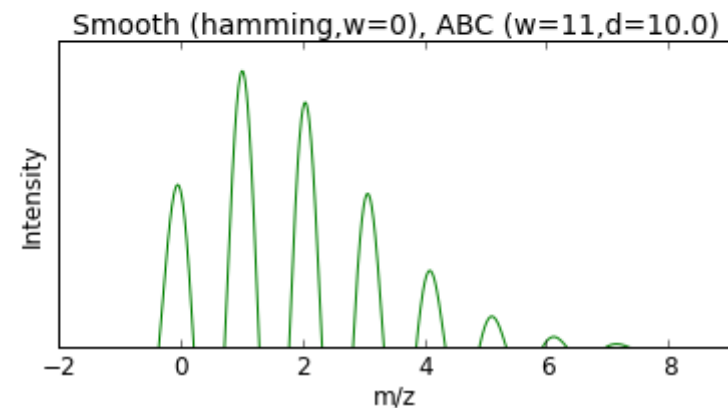
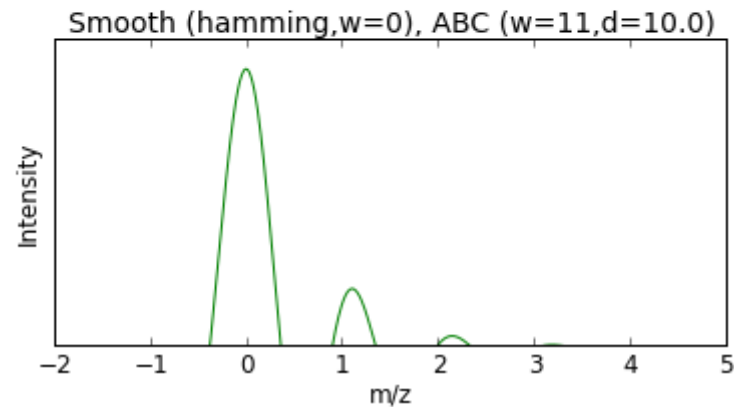
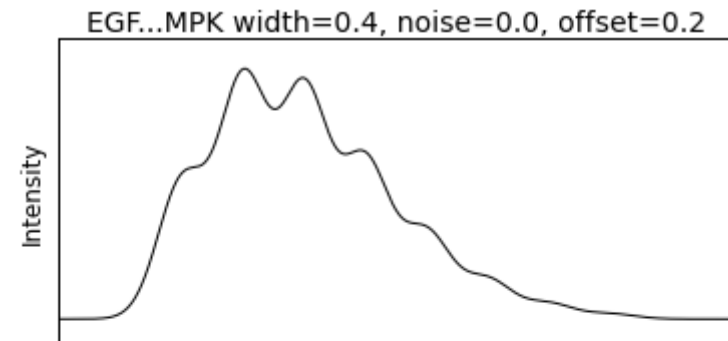
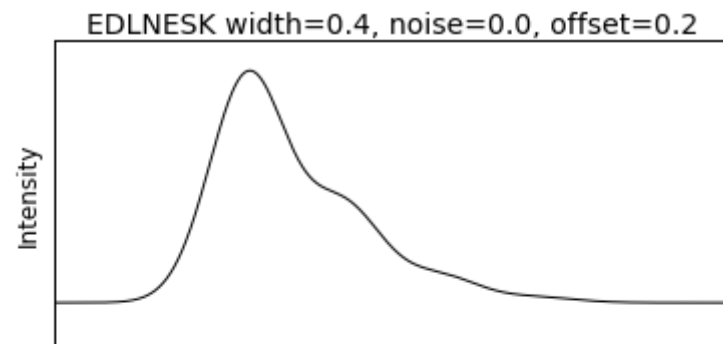
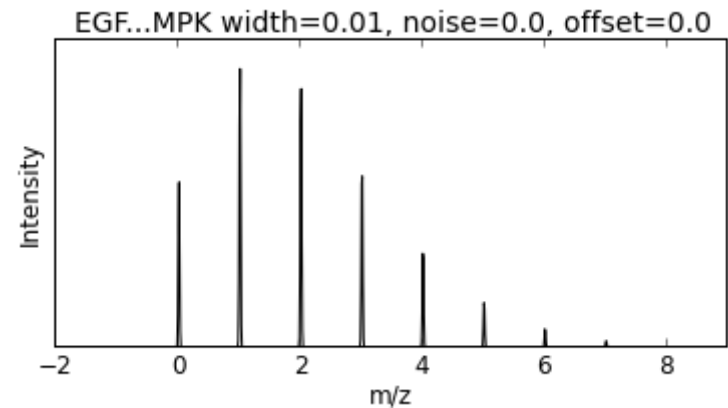
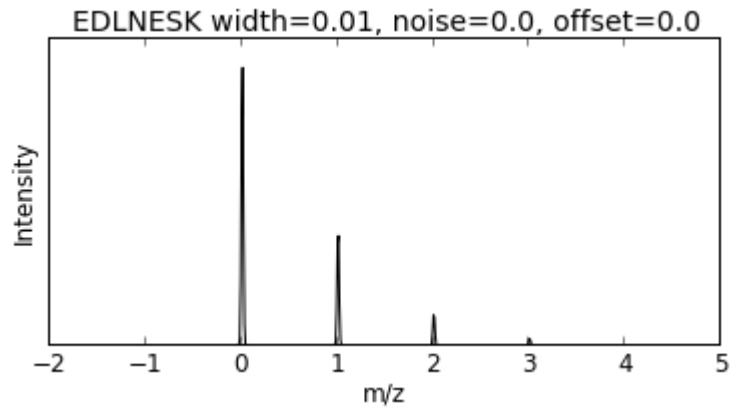
Smooth (hamming,w=11), ABC (w=11,d=0.0)



# Smoothing

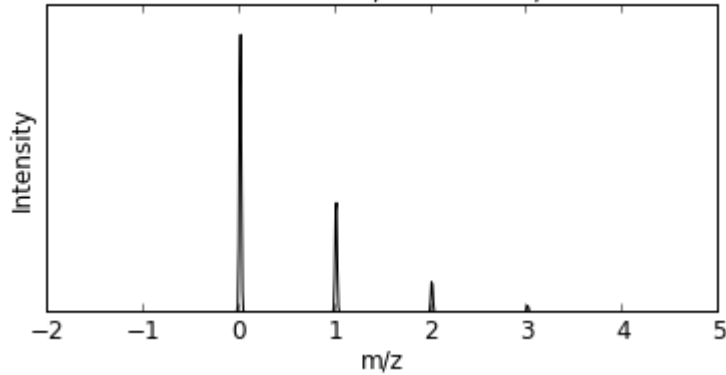


# Adaptive Background Correction

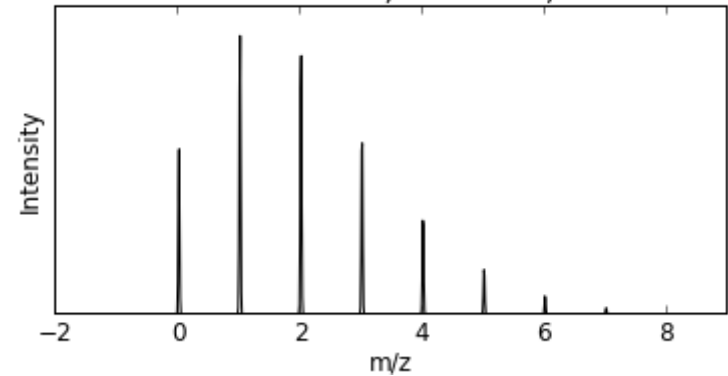


# Smoothing and Adaptive Background Correction

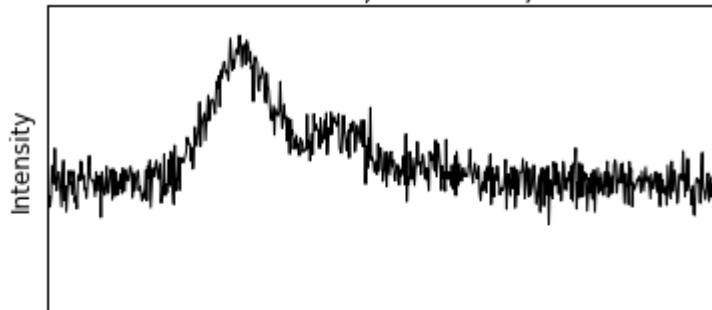
EDLNESK width=0.01, noise=0.0, offset=0.0



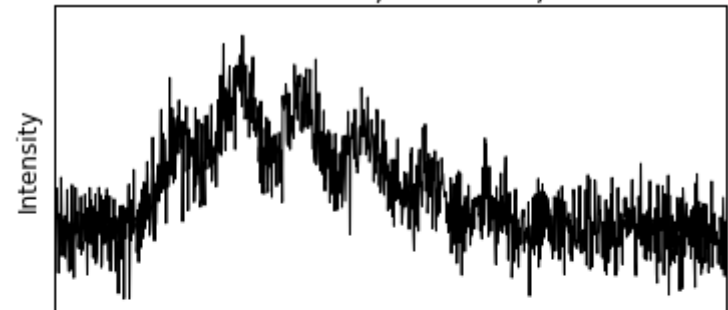
EGF...MPK width=0.01, noise=0.0, offset=0.0



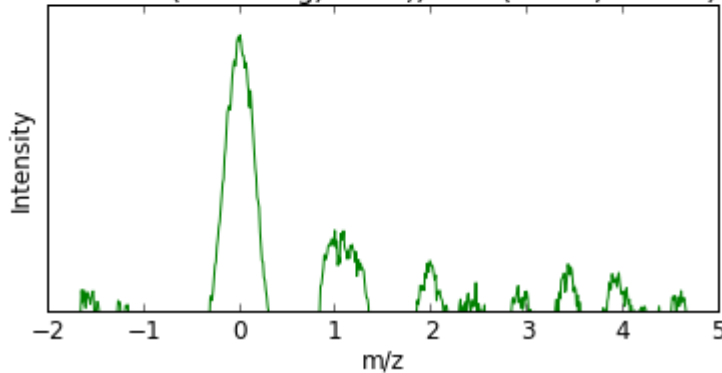
EDLNESK width=0.3, noise=0.1, offset=1.0



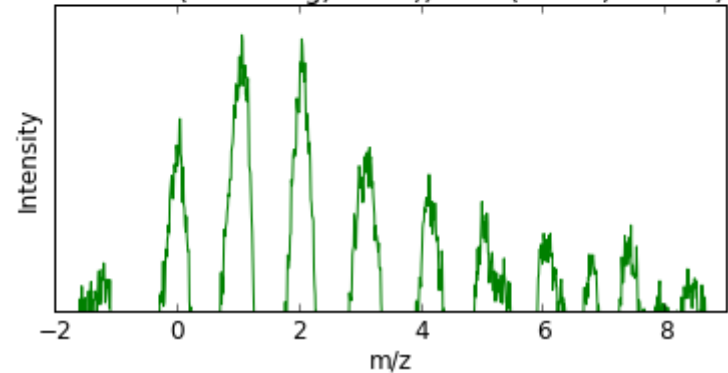
EGF...MPK width=0.3, noise=0.3, offset=1.0



Smooth (hamming,w=31), ABC (w=11,d=10.0)

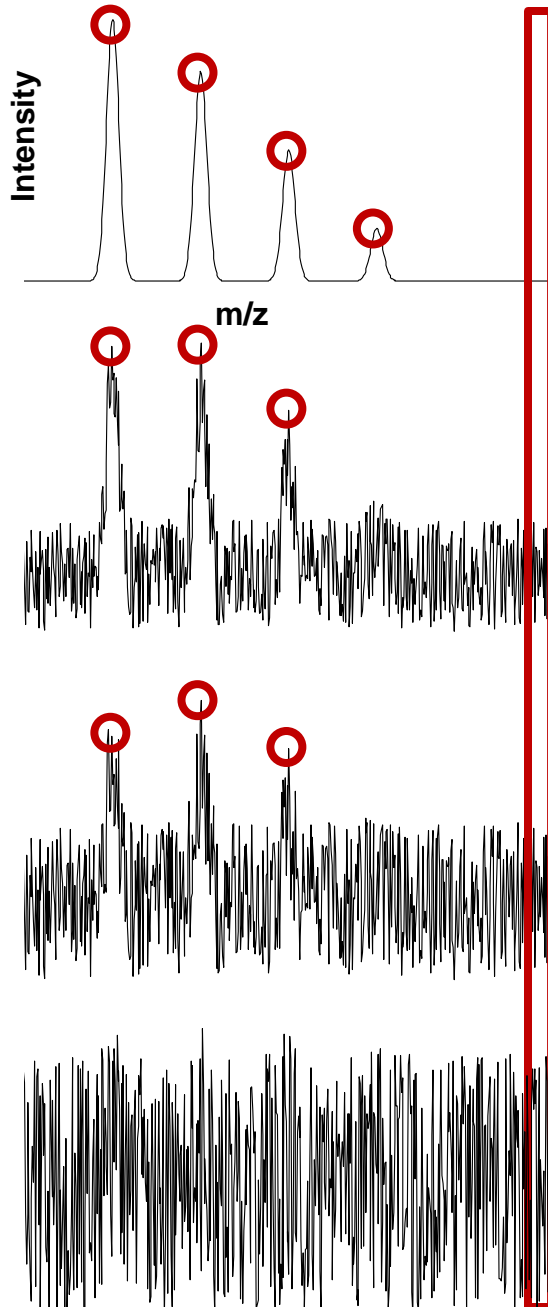


Smooth (hamming,w=31), ABC (w=11,d=10.0)





# Peak Finding



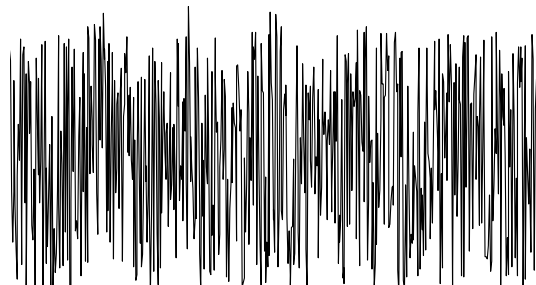
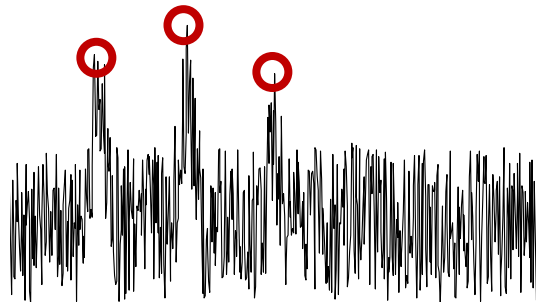
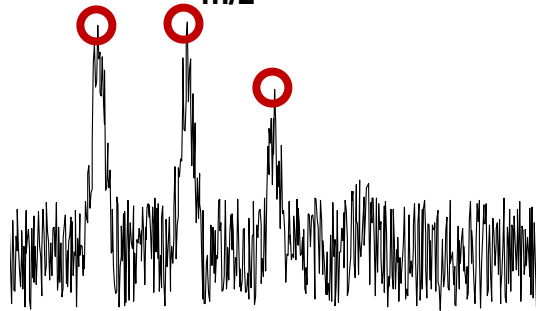
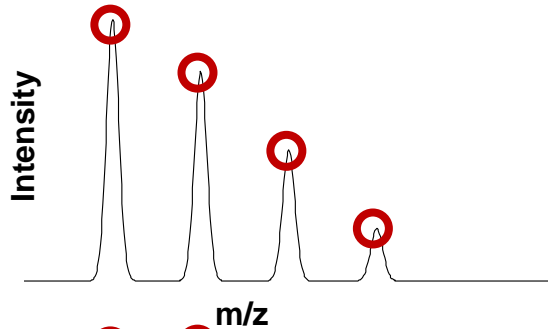
Find maxima of

$$S(I) = \sum_{|k-I| < w/2} I(k)$$

The centroid  $m/z$  of a peak

$$\frac{\sum_{|k-I| < w/2} I(k) \cdot \frac{m}{z}(k)}{\sum_{|k-I| < w/2} I(k)}$$

# Peak Finding



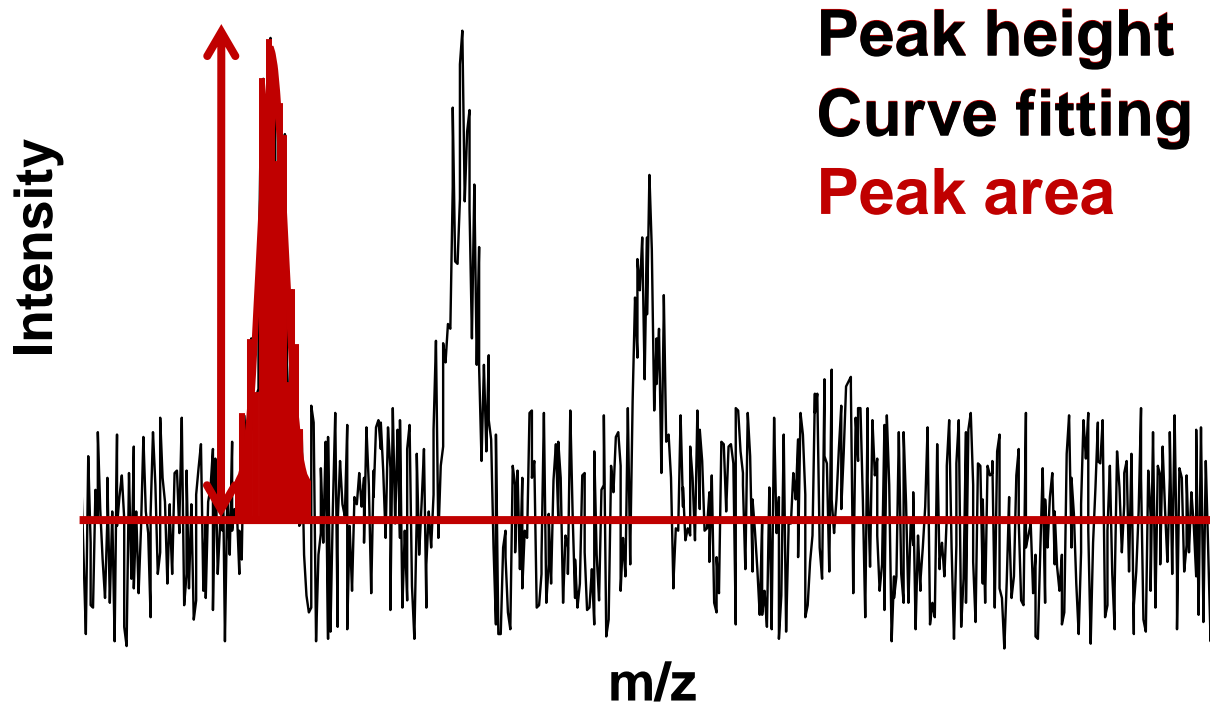
The signal in a peak can be estimated with the RMSD

$$\sqrt{\frac{\sum_{|k-l| < w/2} (I(k) - \langle I \rangle)^2}{w/2}}$$

and the signal-to-noise ratio of a peak can be estimated by dividing the signal with the RMSD of the background

# Estimating peptide quantity

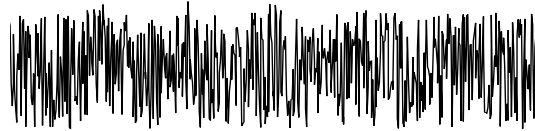
---



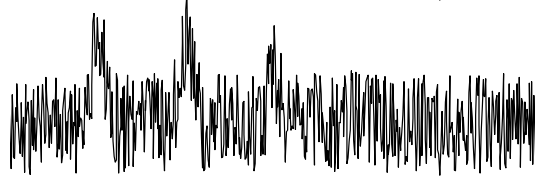
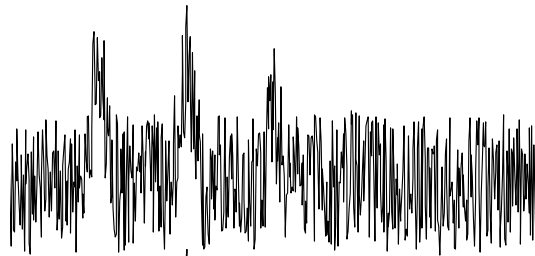
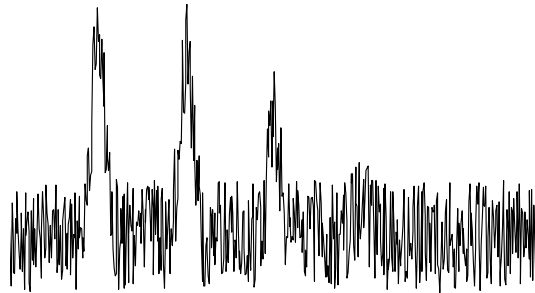
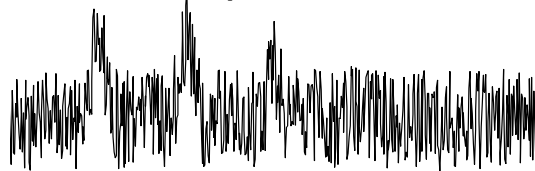
# Time dimension

Time

Intensity



$m/z$



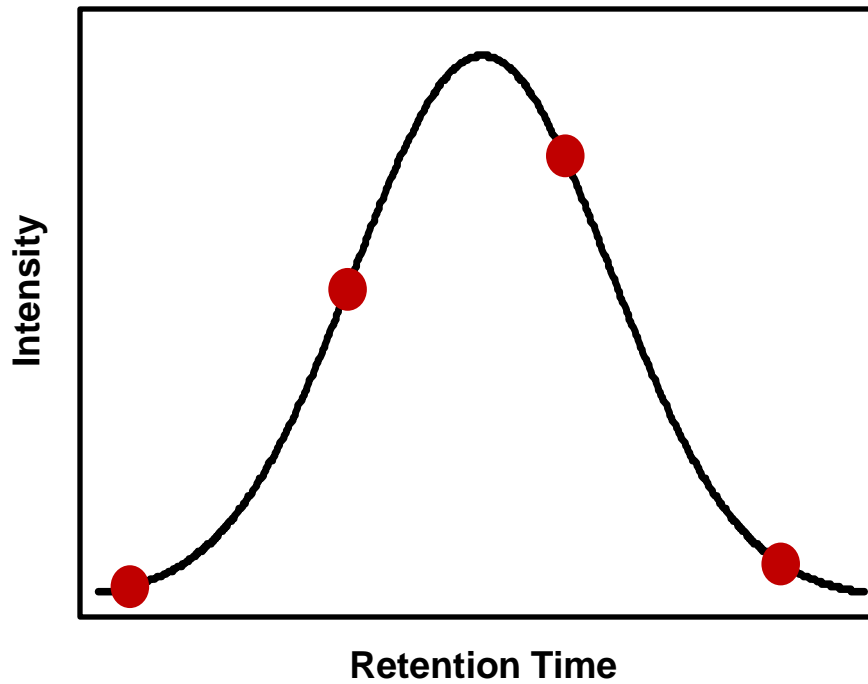
Time



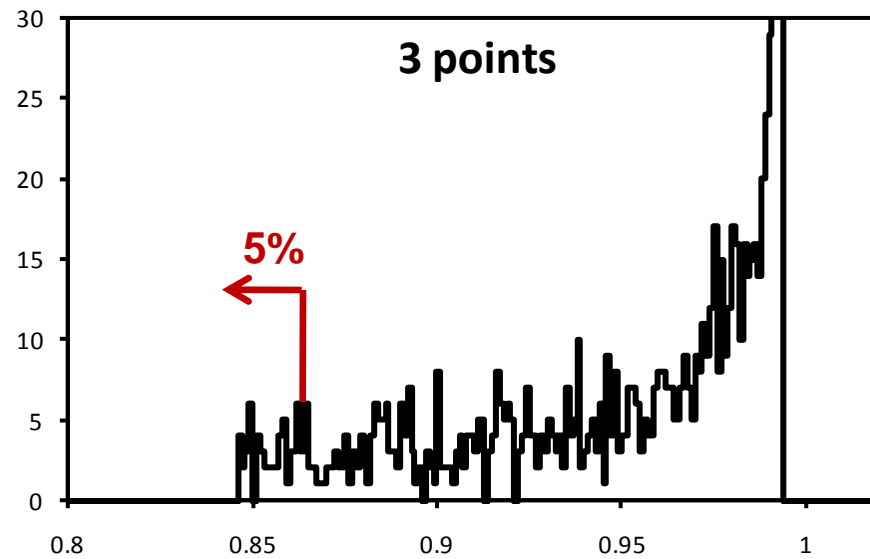
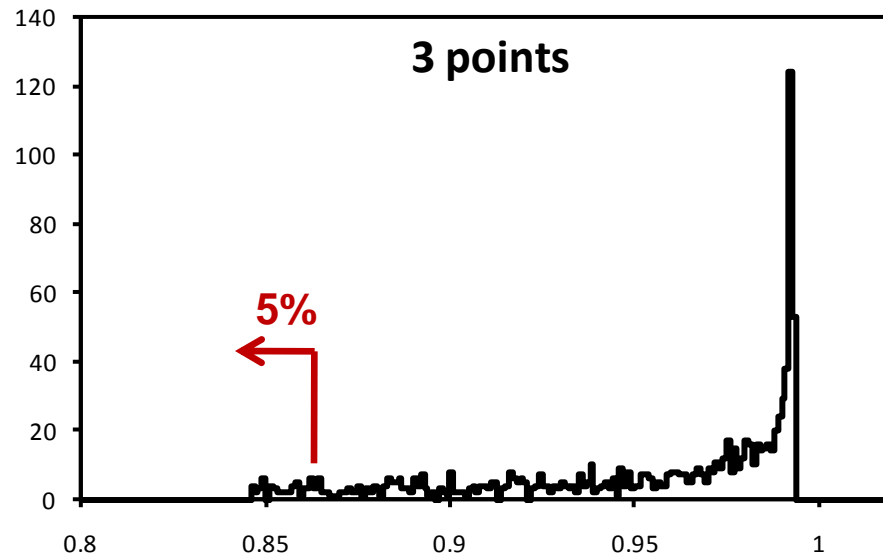
$m/z$

# Sampling

---



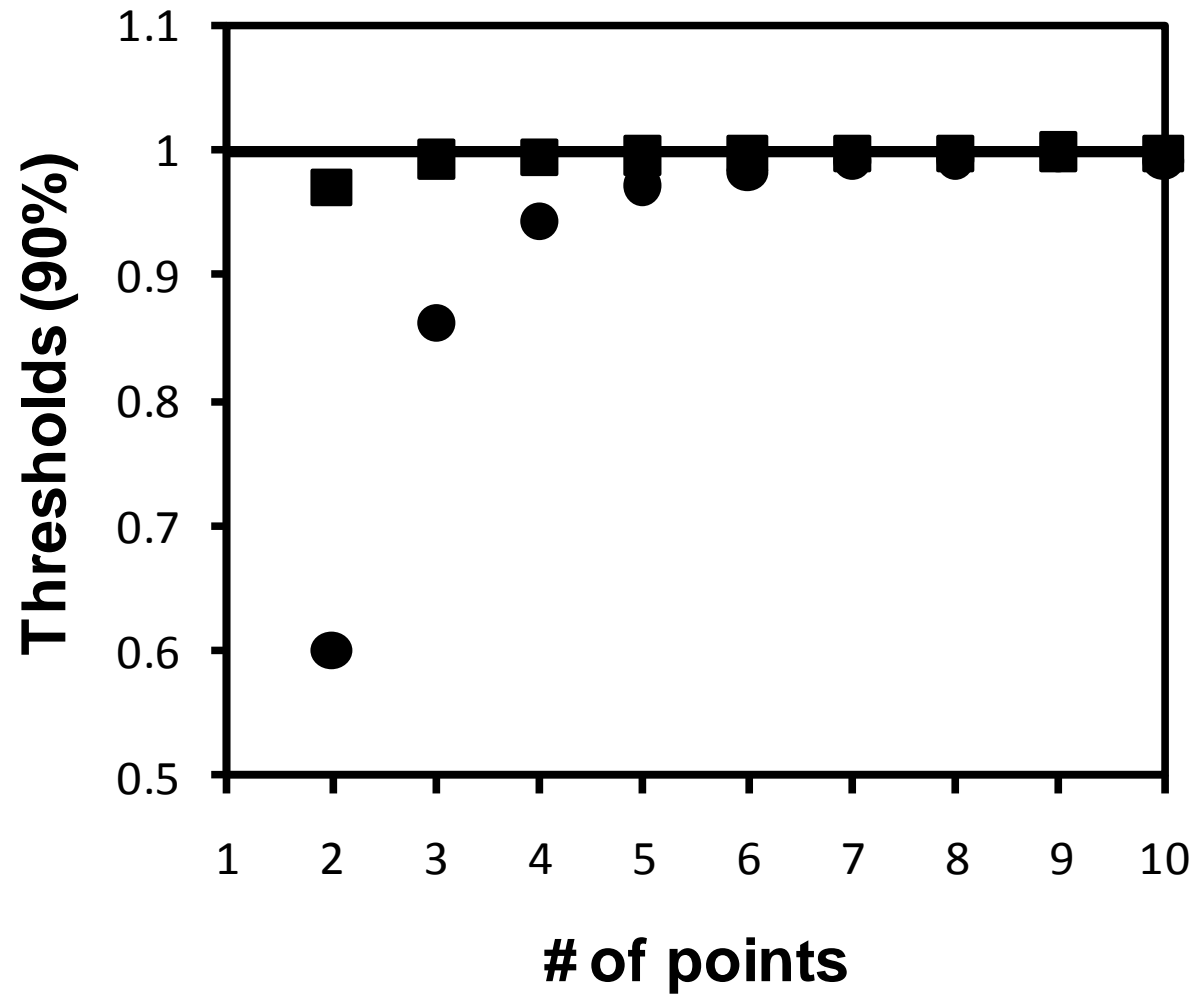
# Sampling



Acquisition time =  $0.05\sigma$

# Sampling

---



# What is the best way to estimate quantity?

Peak height

- resistant to interference
- poor statistics

Peak area

- better statistics
- more sensitive to interference

Curve fitting

- better statistics
- needs to know the peak shape
- slow



# Web Tool

---

[http://10.193.36.101/plot-filter-cgi/plot\\_filter.pl](http://10.193.36.101/plot-filter-cgi/plot_filter.pl) or  
[http://10.193.36.219/plot-filter-cgi/plot\\_filter.pl](http://10.193.36.219/plot-filter-cgi/plot_filter.pl)

peptide:  ▼

Peak width:

Points per m/z unit:

Noise:

Offset:

☐ Apply filters

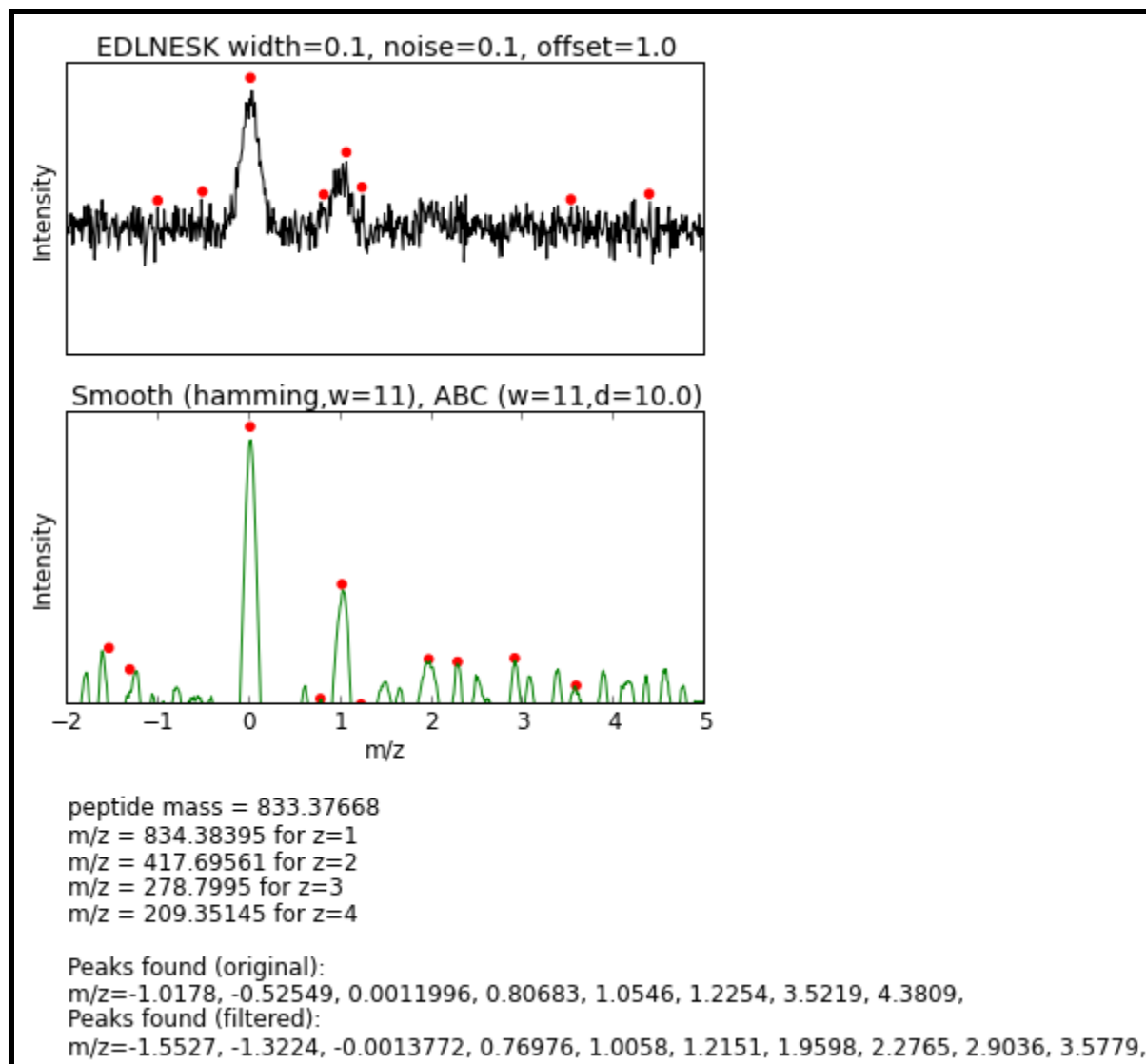
Smoothing:  ▼ width:

Adaptive Background Correction: width:  , strength:

☐ Find peaks

# Web Tool

<http://10.193.36.101> or <http://10.193.36.219>



## Proteomics Informatics -

**Analysis of mass spectra: signal processing, peak finding, and isotope clusters (Week 3)**

---