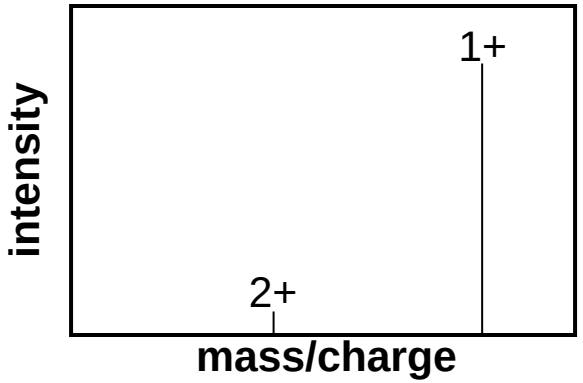
Proteomics Informatics – Analysis of mass spectra: signal processing, peak finding, and isotope clusters (Week 3)

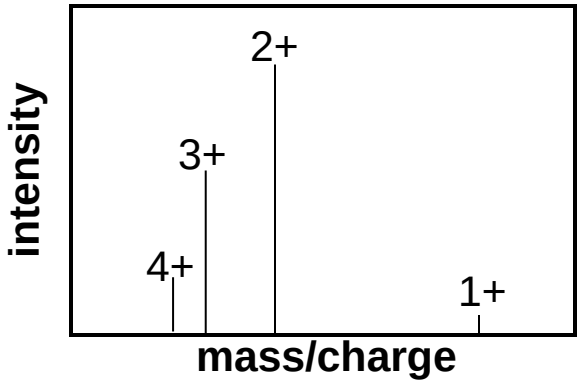
Charge-State Distributions

Peptide

MALDI



ESI

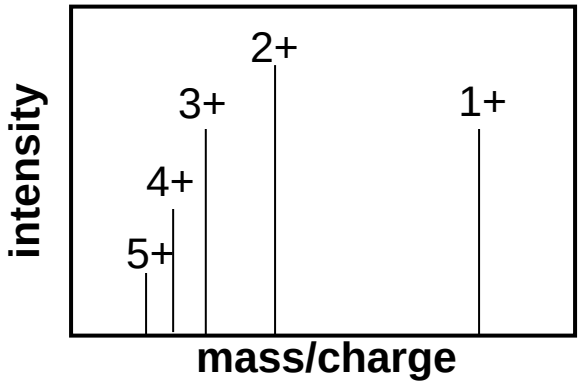


$$\frac{m}{z} = \frac{M + nH}{n}$$

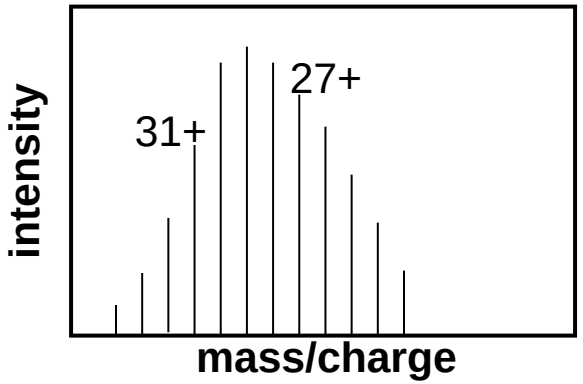
M - molecular mass
n - number of charges
H – mass of a proton

Protein

MALDI



ESI



Charge-State

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

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

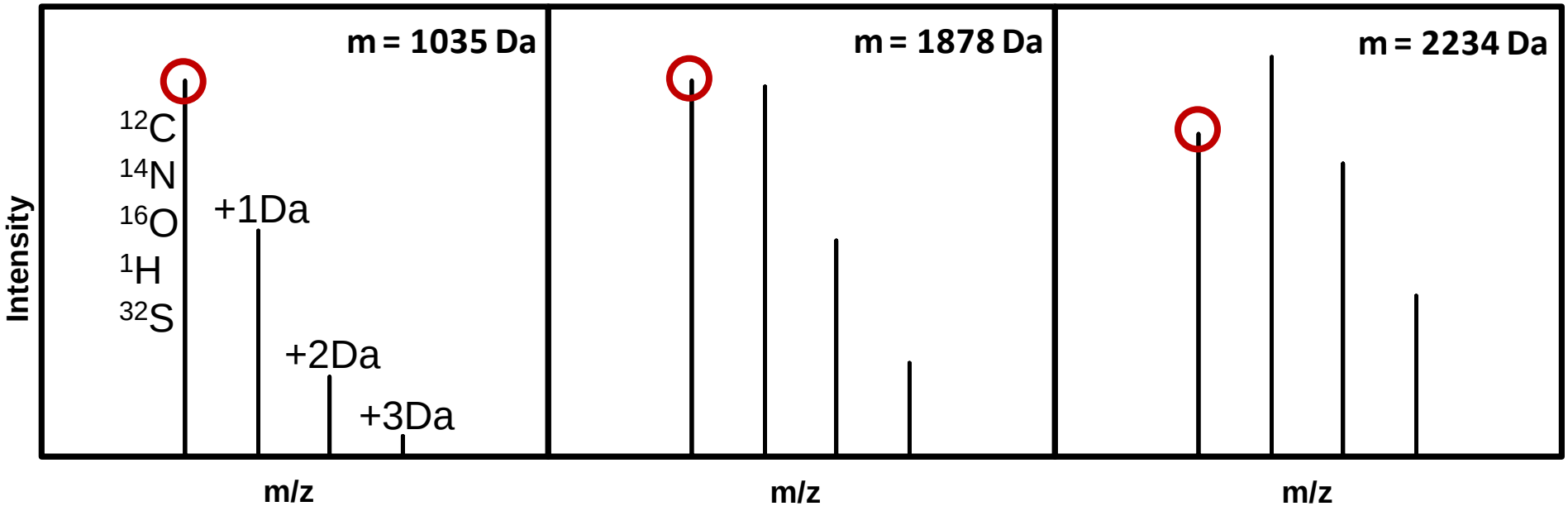
Example:

peptide of mass 898 carrying 1 H^+ = (898 + 1) / 1 = 899 m/z

carrying 2 H^+ = (898 + 2) / 2 = 450 m/z

carrying 3 H^+ = (898 + 3) / 3 = 300.3 m/z

Isotope Distributions



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

Only ^{12}C and ^{13}C:

p=0.0111

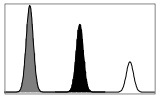
n is the number of C in the peptide

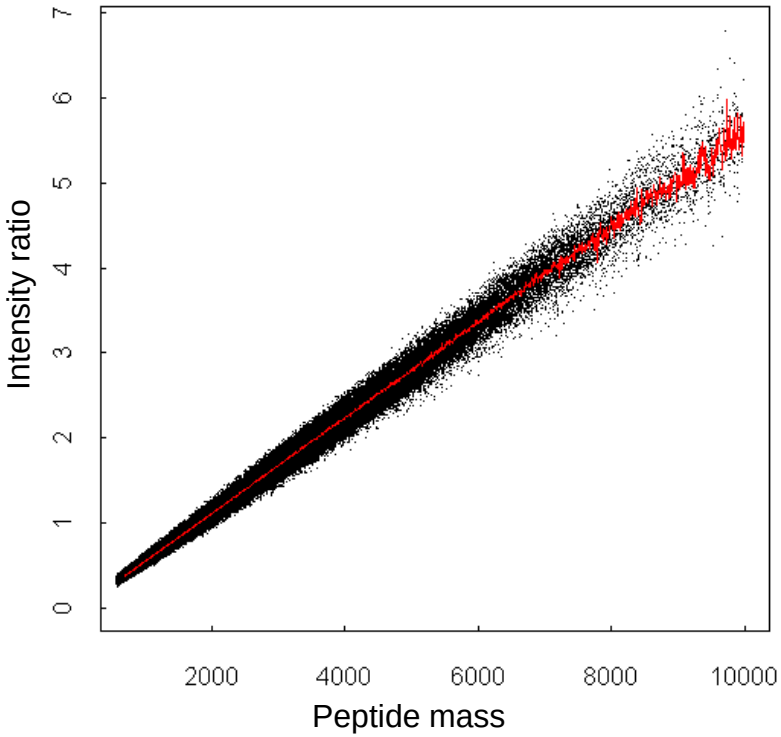
m is the number of ^{13}C in the peptide

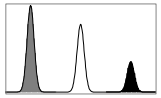
T_m is the relative intensity of the peptide m ^{13}C

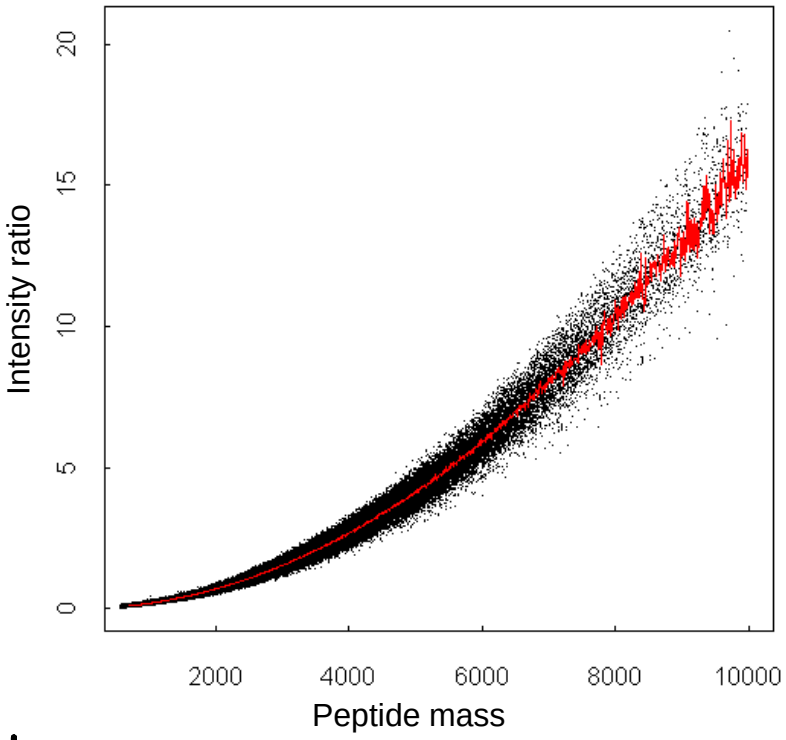
$$T_m = \binom{n}{m} p^m (1-p)^{n-m}$$

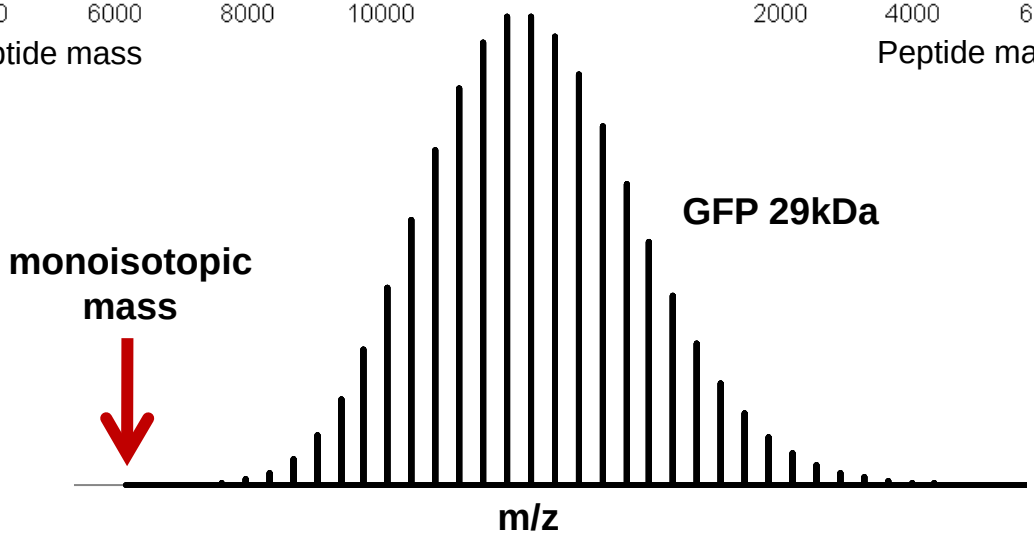
Isotope distributions









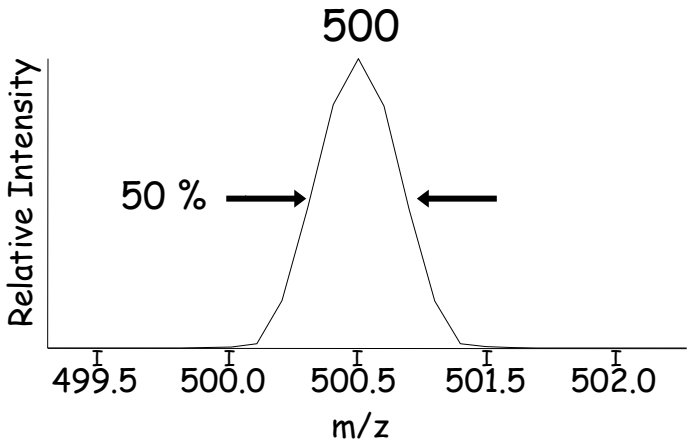


Resolution

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

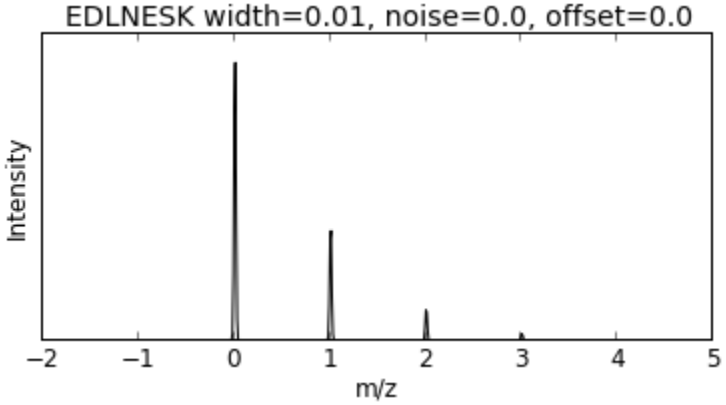
Resolution = minimum peak separation, ΔM,
which allows to distinguish two ion species

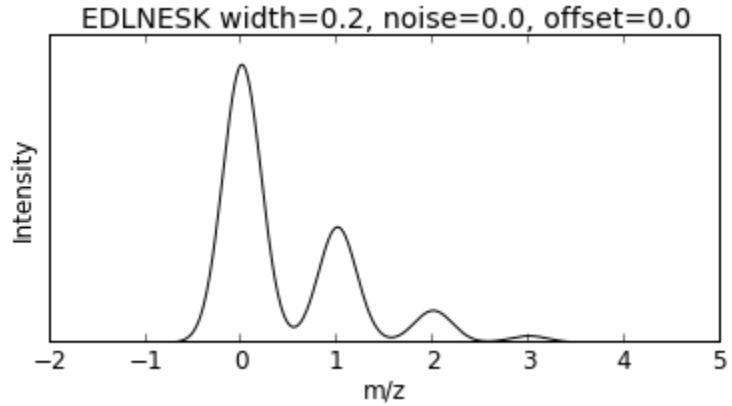
ΔM = full width at half maximum
(FWHM)

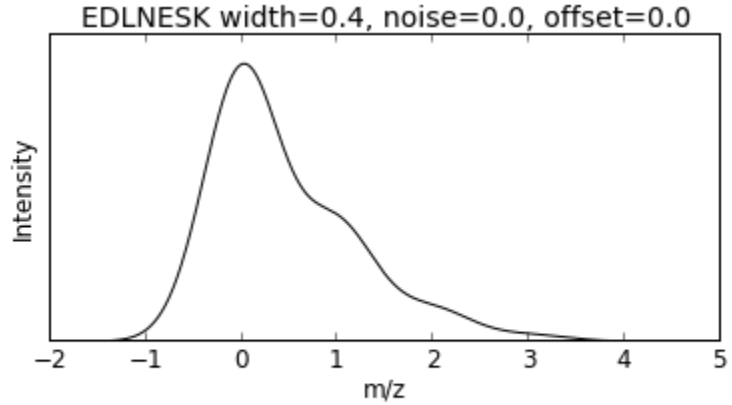


Resolution = $M/\Delta M$ = 500/0.5 = 1000

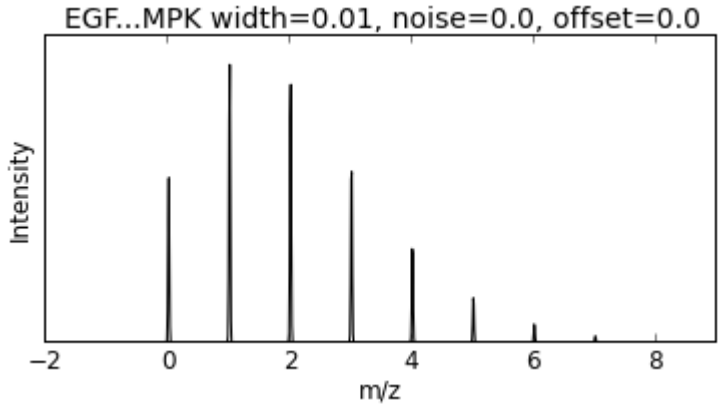
Resolution

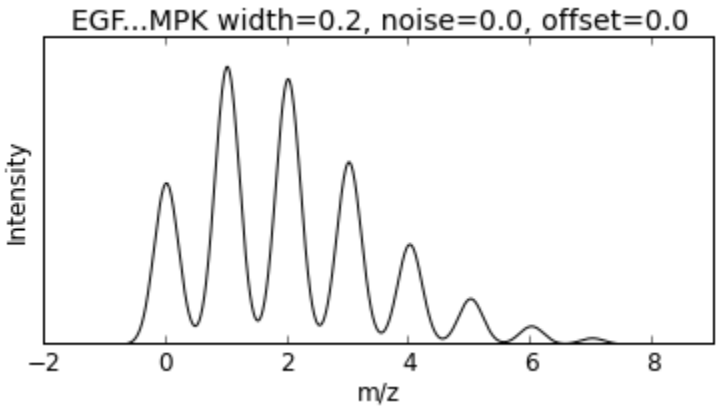


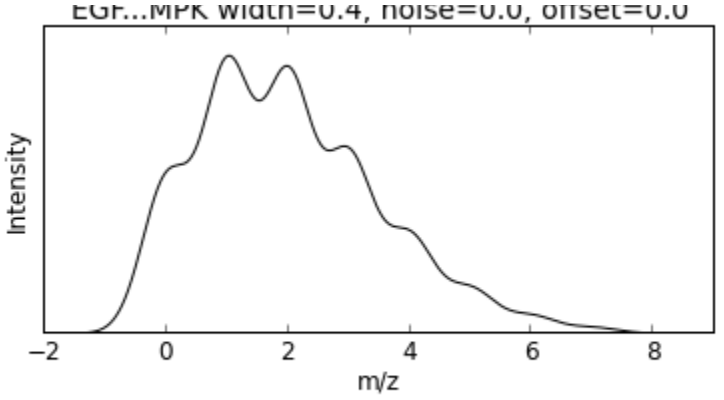




peptide mass = 833.37668
m/z = 834.38395 for z=1







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

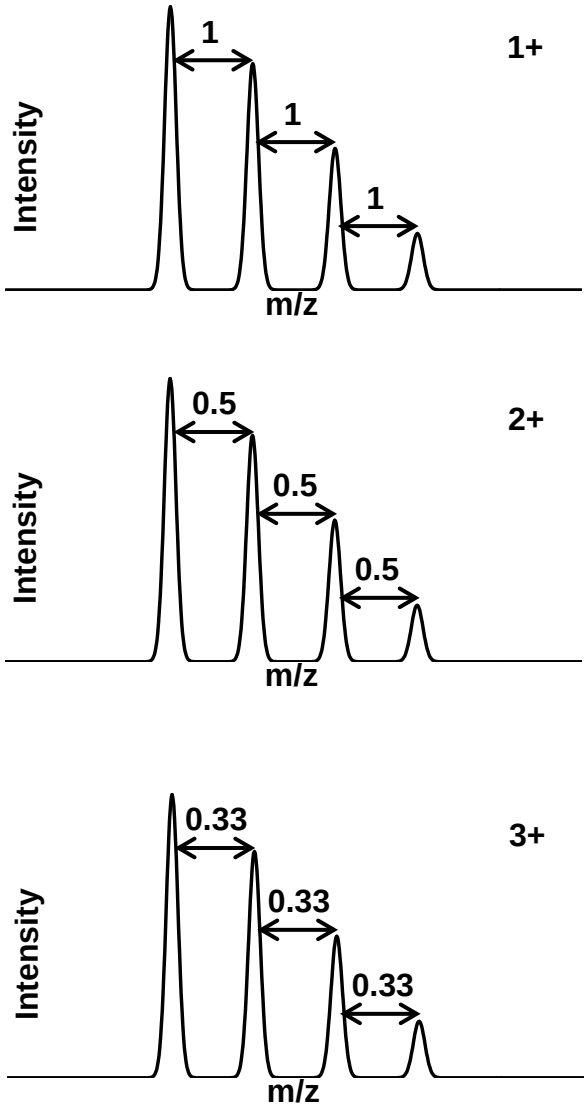
Resolution

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

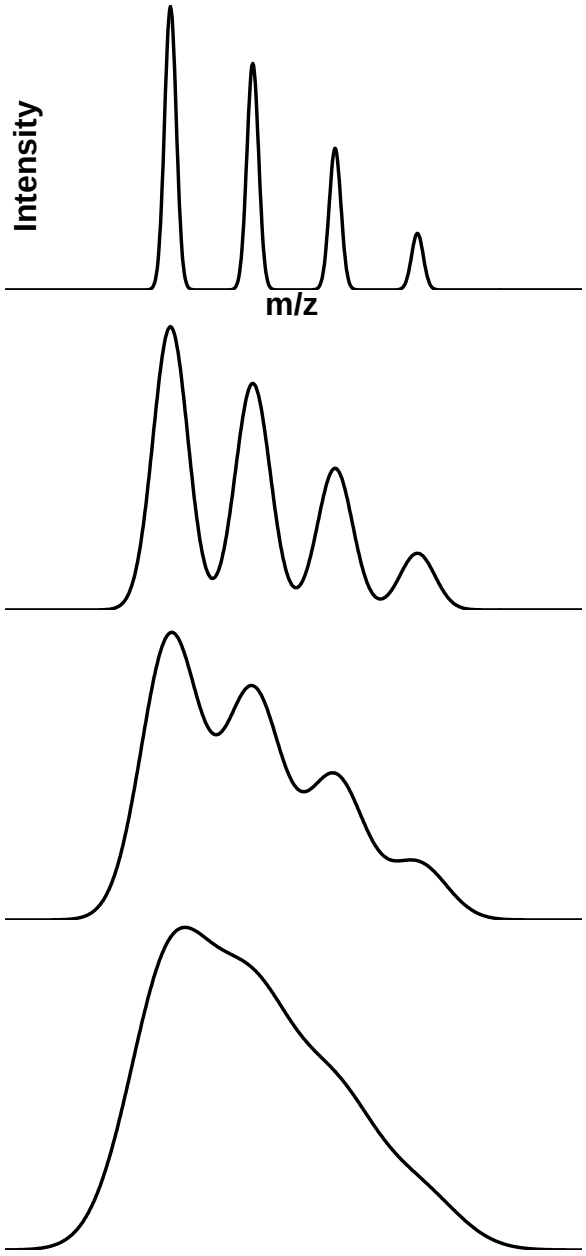
- What resolution do we need to differentiate a 1600 Da peptide that carries either an acetylation (+ 42.0100) or trimethylation (42.0464)?

- R = 1600/0.0364 = 43,956

Isotope Clusters and Charge State



Isotope Clusters and Charge State



Possible to Determine Charge?

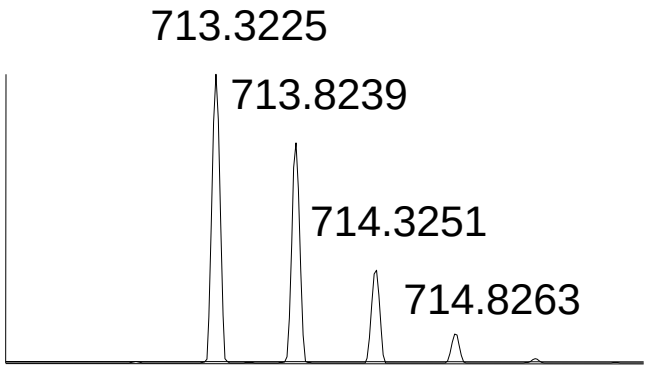
Yes

Yes

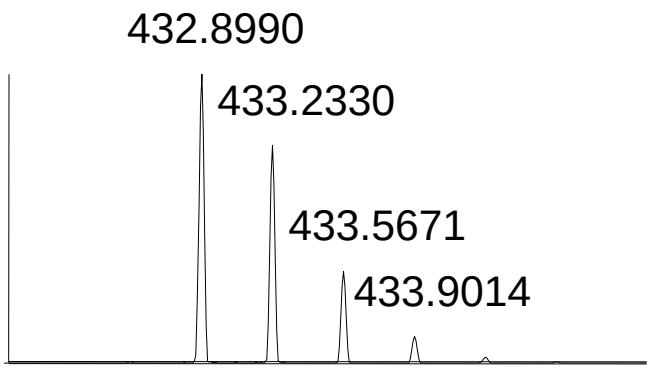
Maybe

No

What is the Charge State?

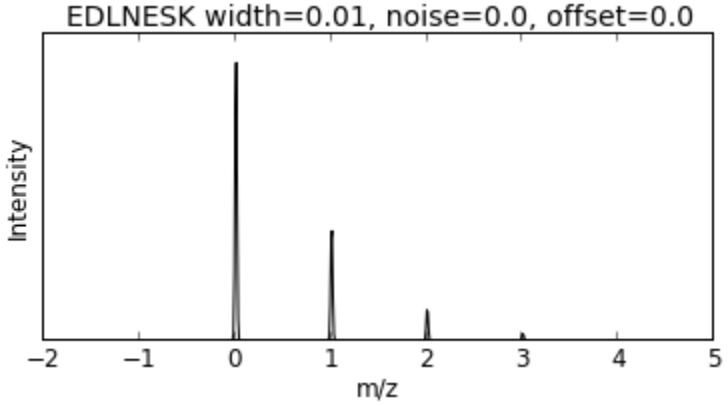


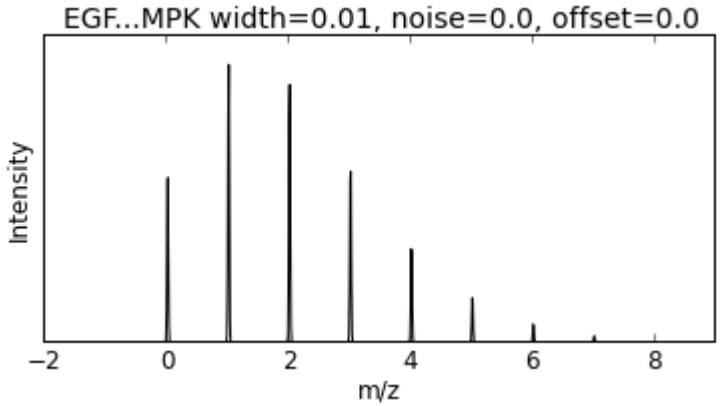
Δ between the isotopes is 0.5 Da

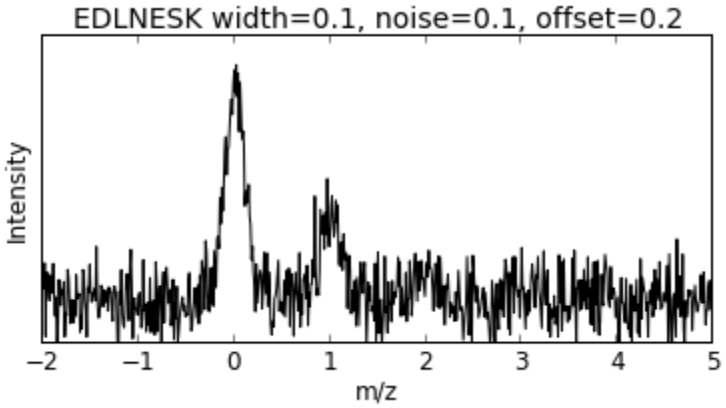


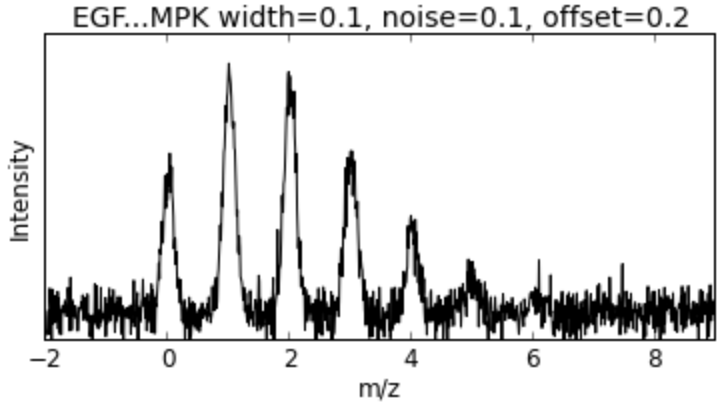
Δ between the isotopes is 0.33 Da

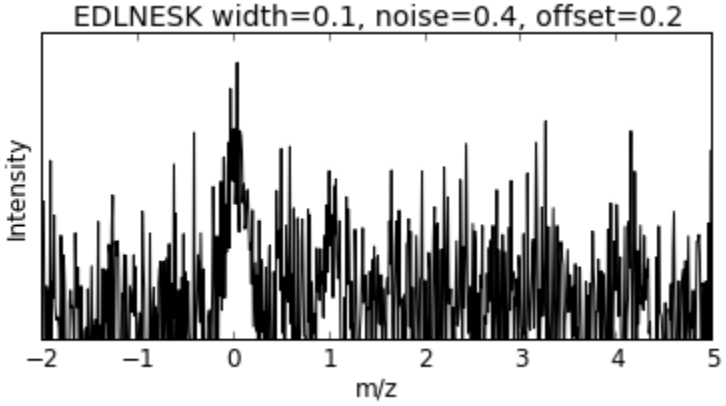
Noise

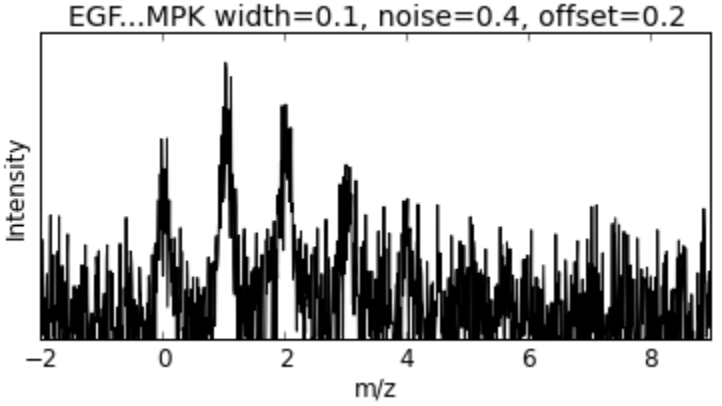

EDLNESK width=0.01, noise=0.0, offset=0.0
Intensity
−2 −1 0 1 2 3 4 5
m/z


EGF...MPK width=0.01, noise=0.0, offset=0.0
Intensity
−2 0 2 4 6 8
m/z

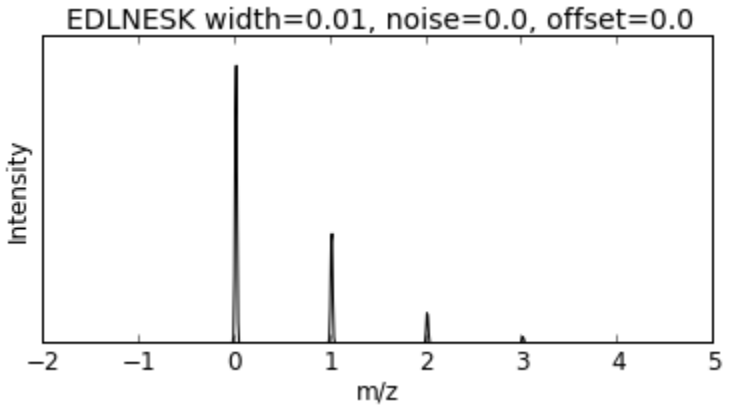

EDLNESK width=0.1, noise=0.1, offset=0.2
Intensity
−2 −1 0 1 2 3 4 5
m/z

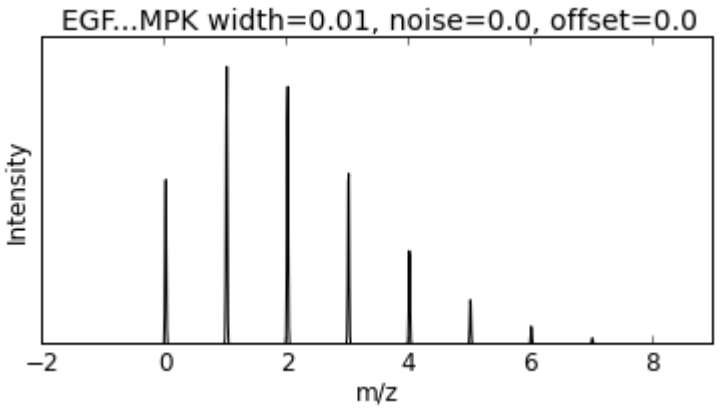

EGF...MPK width=0.1, noise=0.1, offset=0.2
Intensity
−2 0 2 4 6 8
m/z

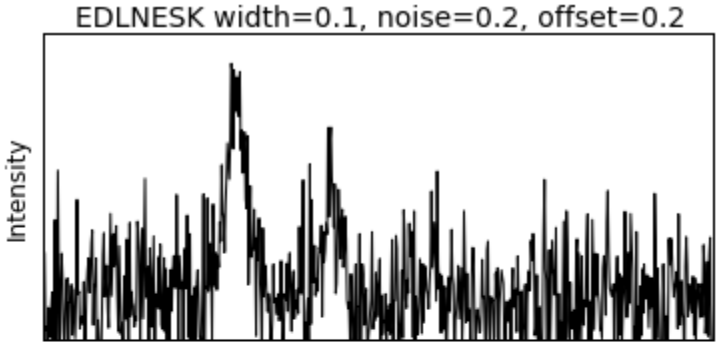

EDLNESK width=0.1, noise=0.4, offset=0.2
Intensity
−2 −1 0 1 2 3 4 5
m/z

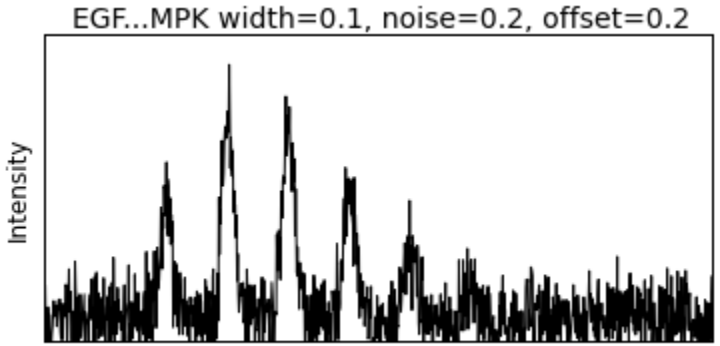

EGF...MPK width=0.1, noise=0.4, offset=0.2
Intensity
−2 0 2 4 6 8
m/z

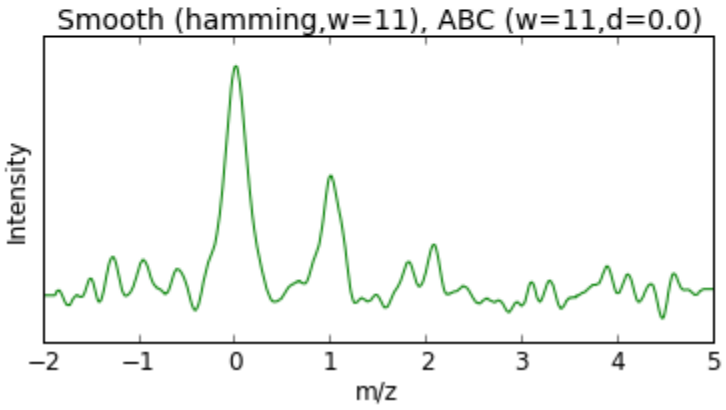
Smoothing

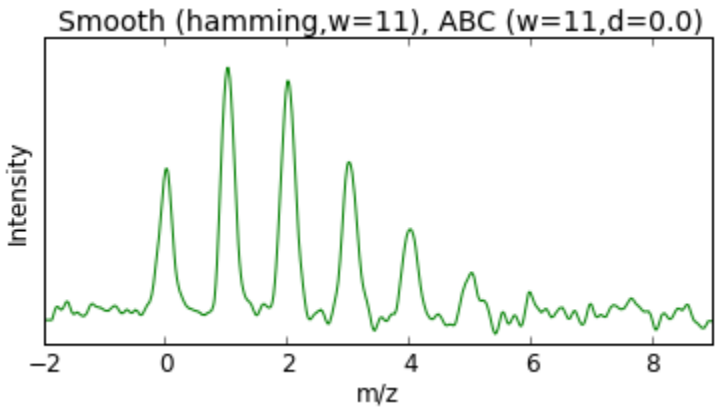

EDLNESK width=0.01, noise=0.0, offset=0.0
Intensity
−2
−1
0
1
2
3
4
5
m/z


EGF...MPK width=0.01, noise=0.0, offset=0.0
Intensity
−2
0
2
4
6
8
m/z

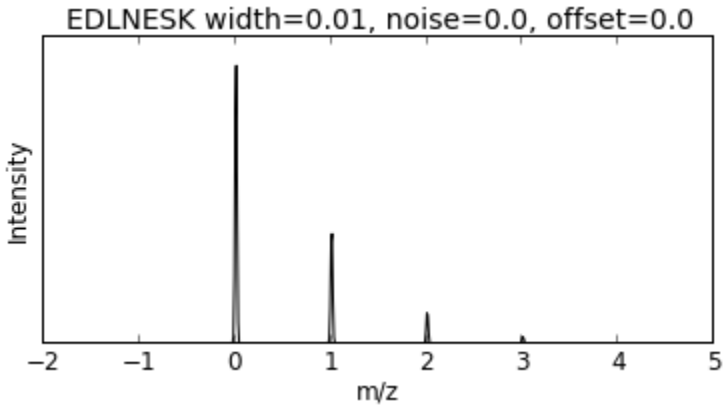

EDLNESK width=0.1, noise=0.2, offset=0.2
Intensity

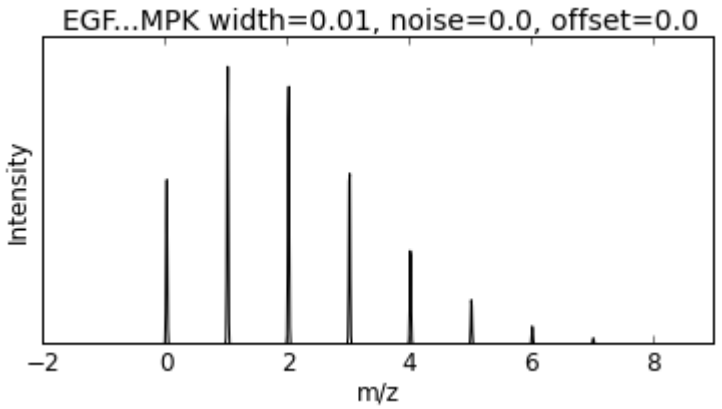

EGF...MPK width=0.1, noise=0.2, offset=0.2
Intensity

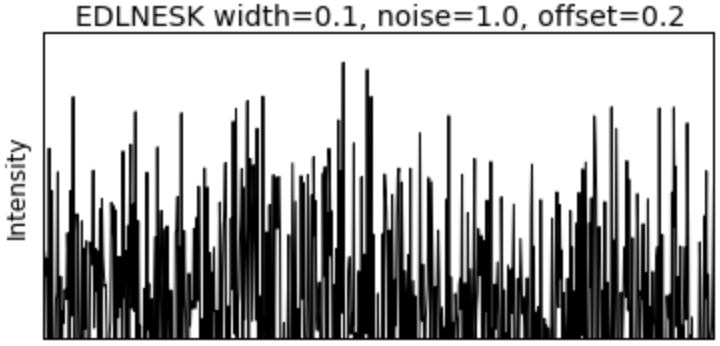

Smooth (hamming,w=11), ABC (w=11,d=0.0)
Intensity
−2
−1
0
1
2
3
4
5
m/z

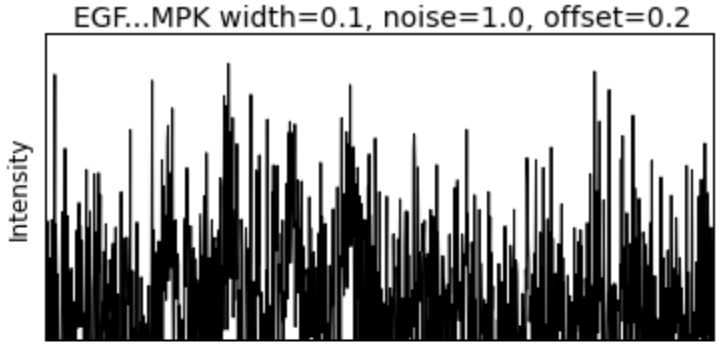

Smooth (hamming,w=11), ABC (w=11,d=0.0)
Intensity
−2
0
2
4
6
8
m/z

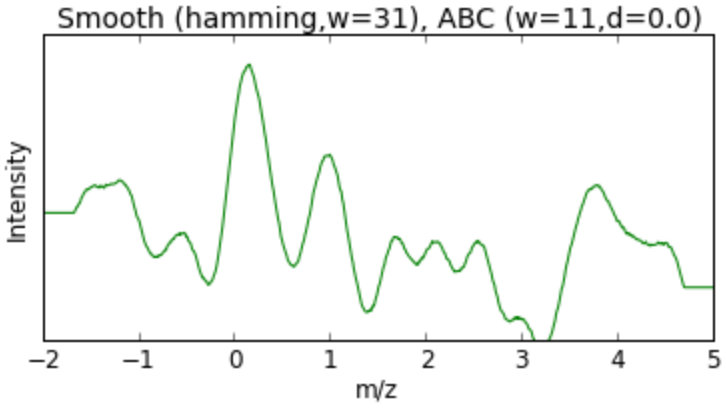
Smoothing

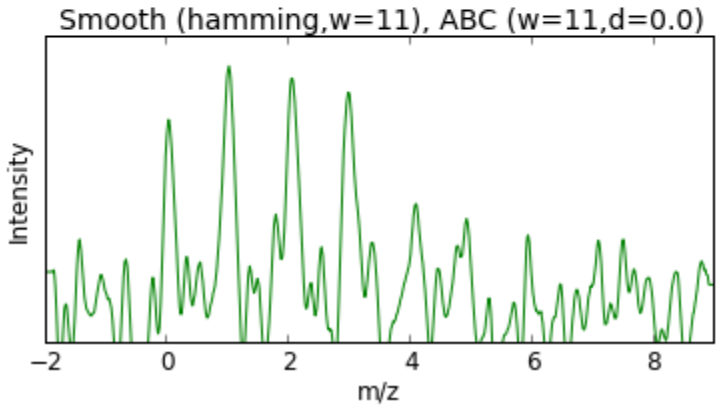

EDLNESK width=0.01, noise=0.0, offset=0.0
Intensity
−2
−1
0
1
2
3
4
5
m/z


EGF...MPK width=0.01, noise=0.0, offset=0.0
Intensity
−2
0
2
4
6
8
m/z


EDLNESK width=0.1, noise=1.0, offset=0.2
Intensity


EGF...MPK width=0.1, noise=1.0, offset=0.2
Intensity

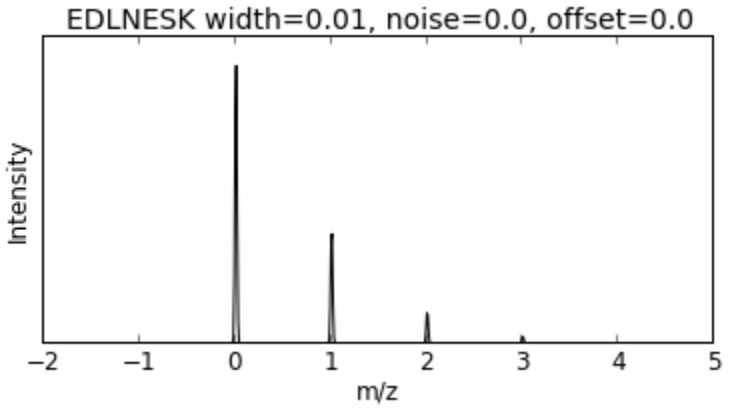

Smooth (hamming,w=31), ABC (w=11,d=0.0)
Intensity
−2
−1
0
1
2
3
4
5
m/z

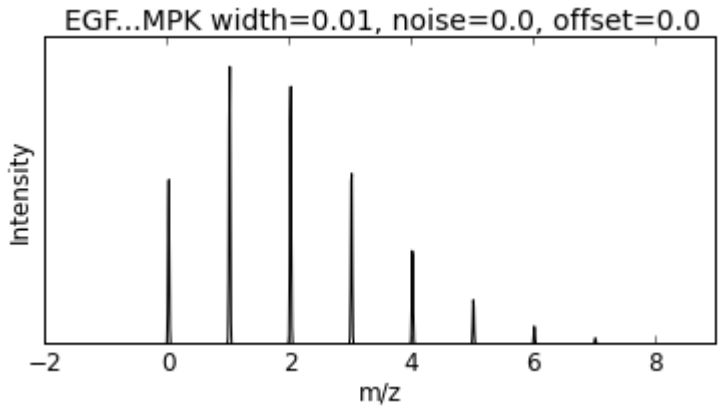

Smooth (hamming,w=11), ABC (w=11,d=0.0)
Intensity
−2
0
2
4
6
8
m/z

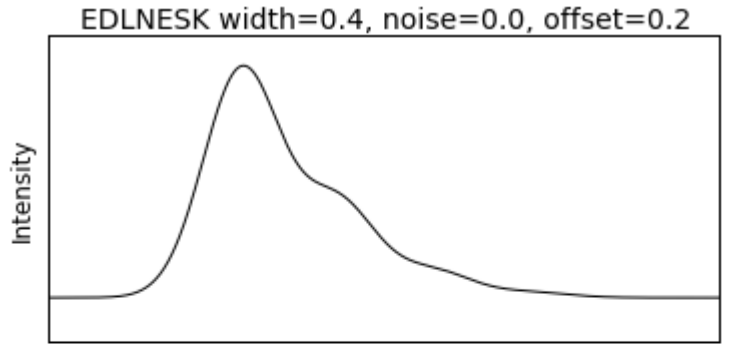
Adaptive Background Correction (Unsharp masking)

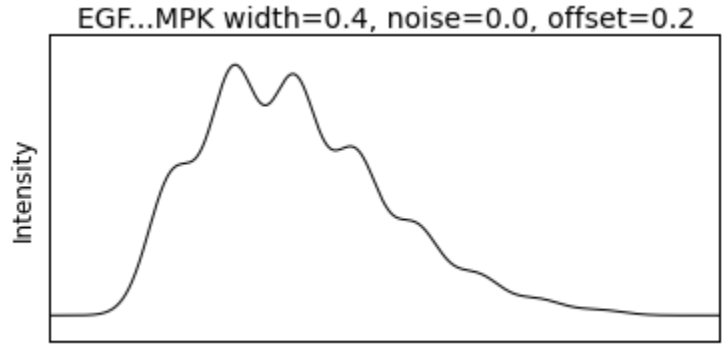
$$I'(l,d,w) = \frac{d}{2w+1} \sum_{k=l-w}^{k=l+w} I(k)$$

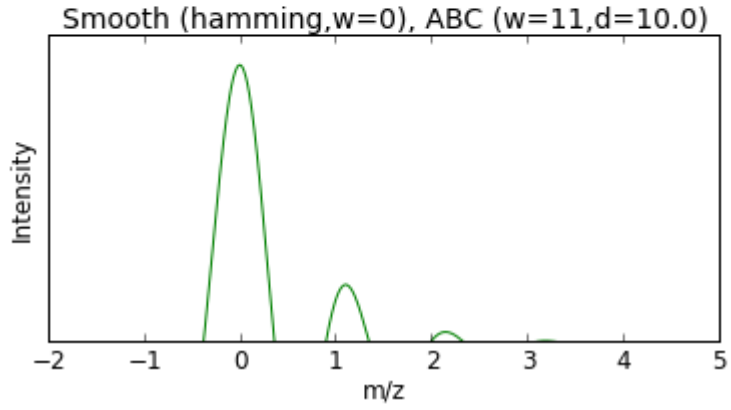
Adaptive Background Correction

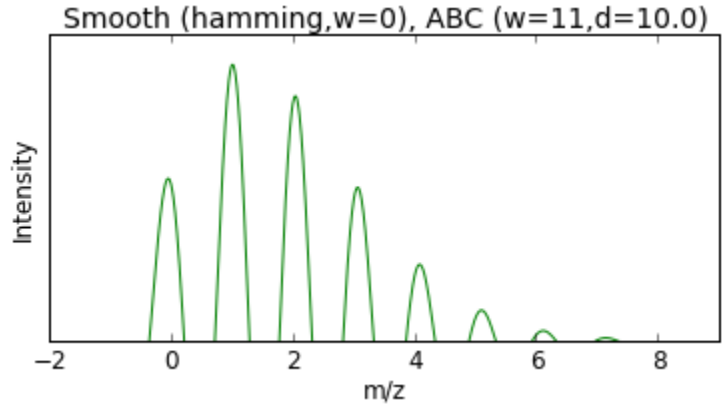




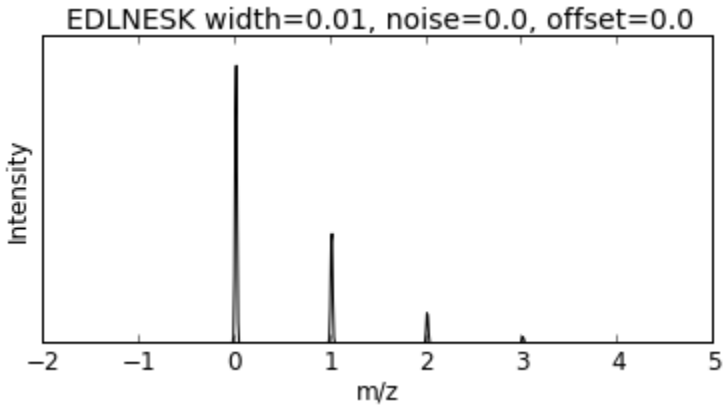


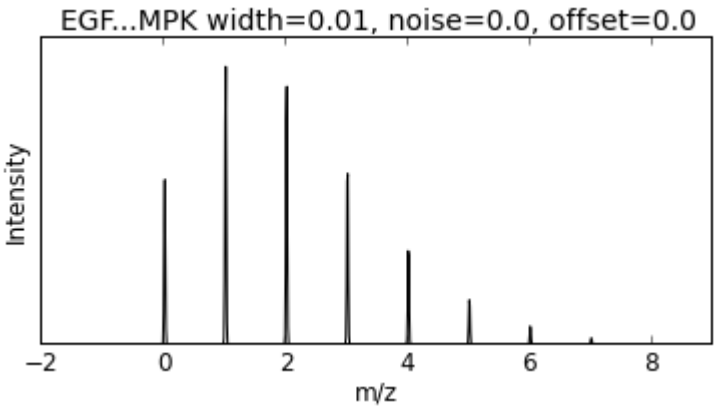


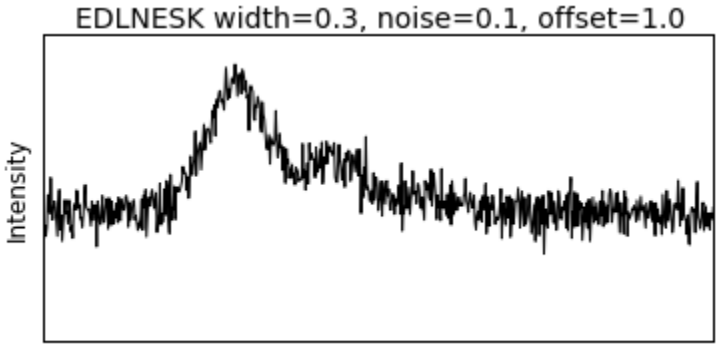


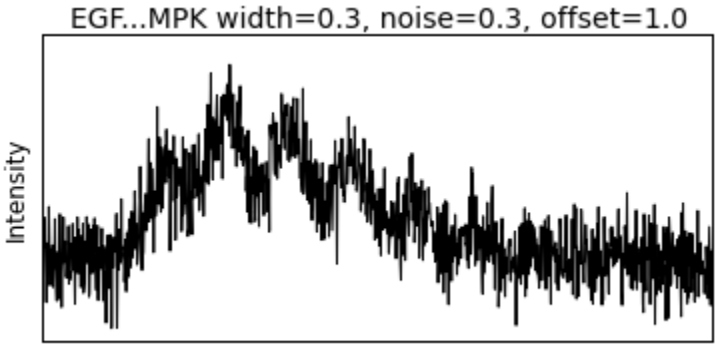


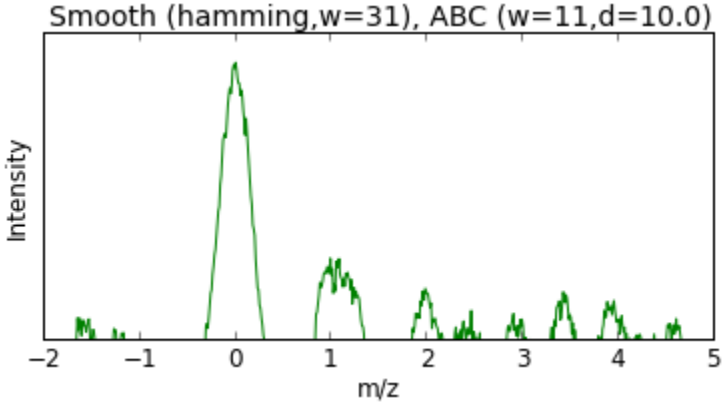
Smoothing and Adaptive Background Correction

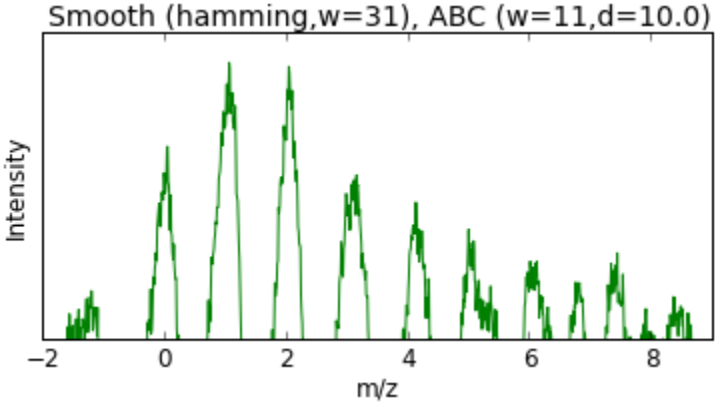




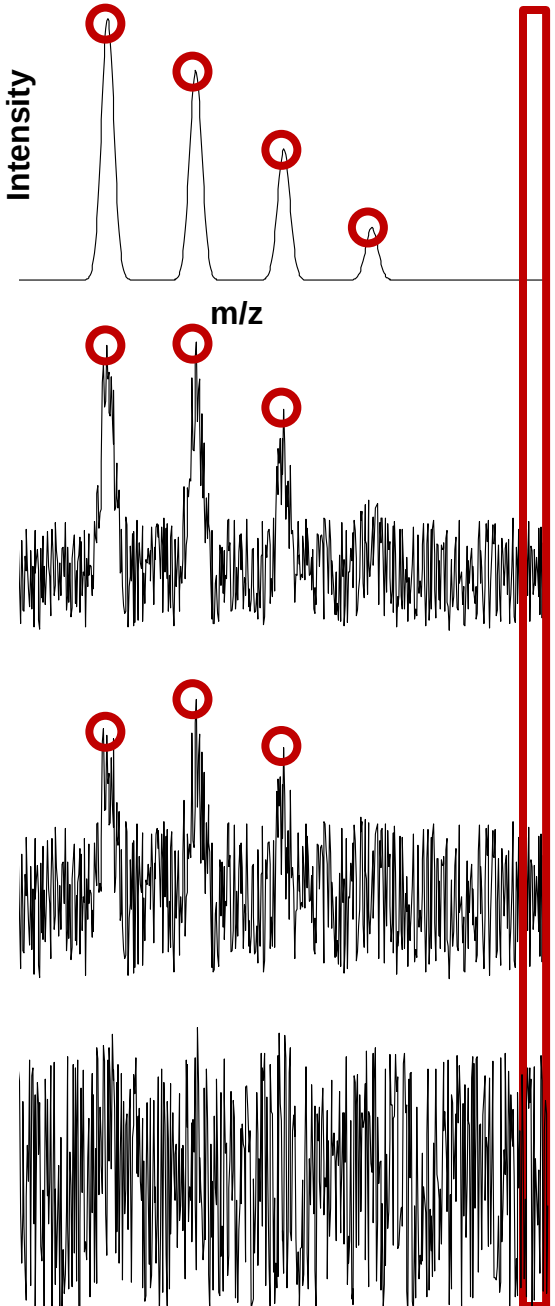








Peak Finding



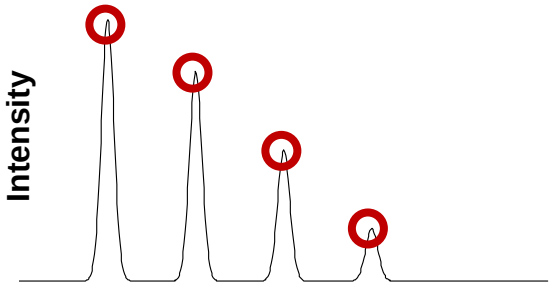
Find maxima of

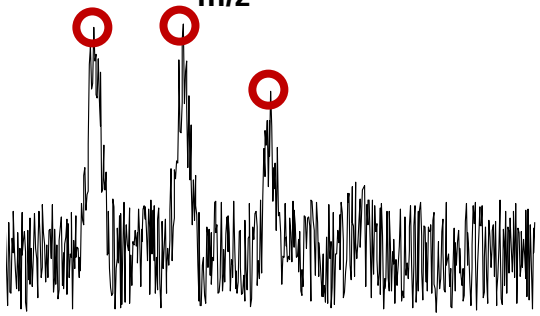
$$S(l) = \sum_{k=l-w}^{k=l+w} I(k)$$

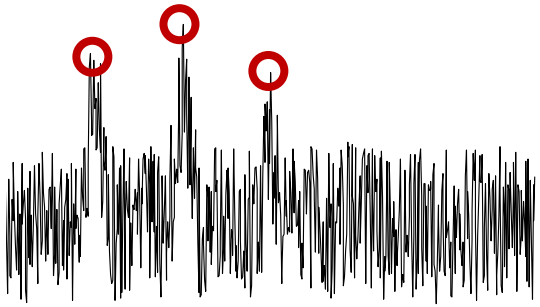
The centroid m/z of a peak

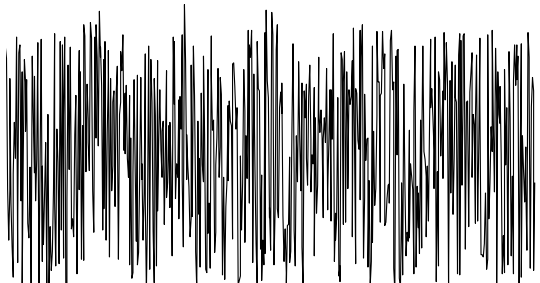
$$\frac{\sum_{k=l-w}^{k=l+w} I(k) \cdot \frac{m}{z}(k)}{\sum_{k=l-w}^{k=l+w} I(k)}$$

Peak Finding







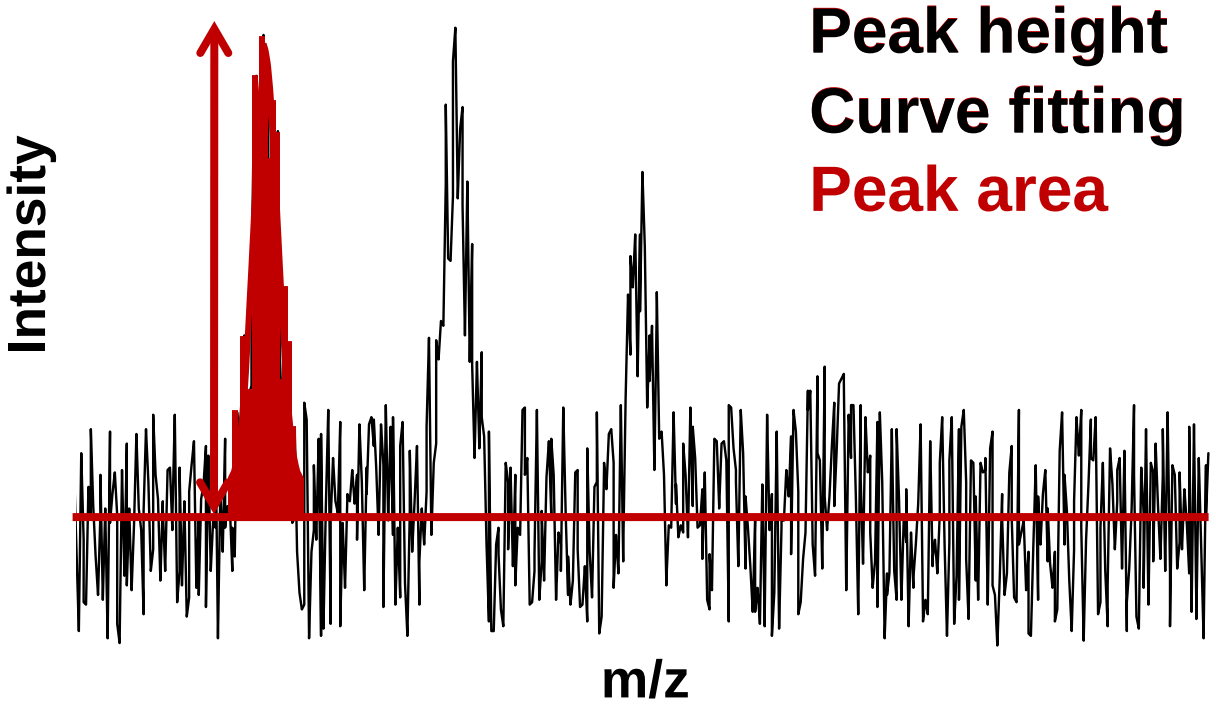


The signal in a peak can be estimated with the RMSD

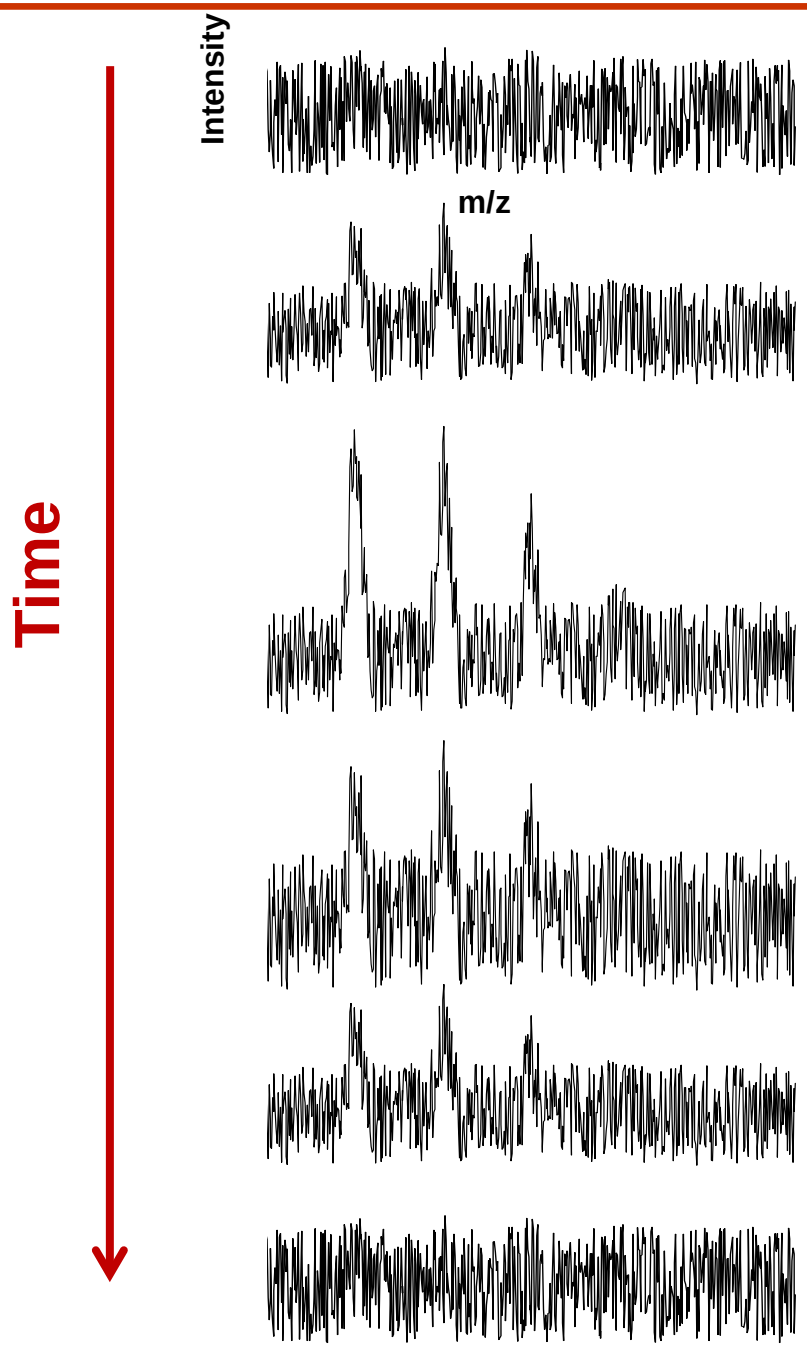
$$\sqrt{\frac{\sum_{|k-l|<w/2} \left(I(k)-<I>\right)^2}{w/2}}$$

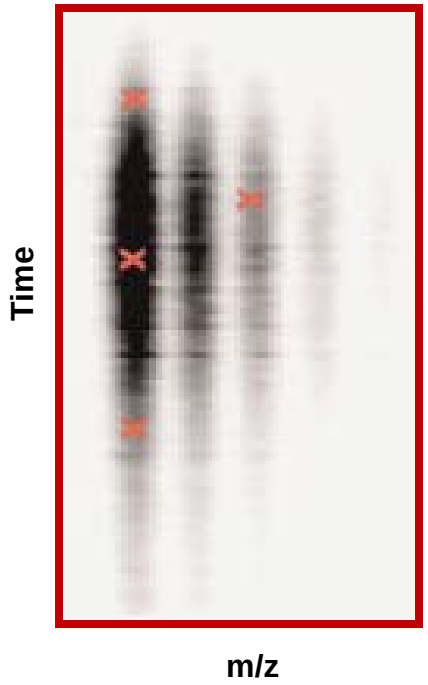
and the signal-to-noise ratio of a peak can be estimated by dividing the signal with the RMSD of the background

Estimating peptide quantity

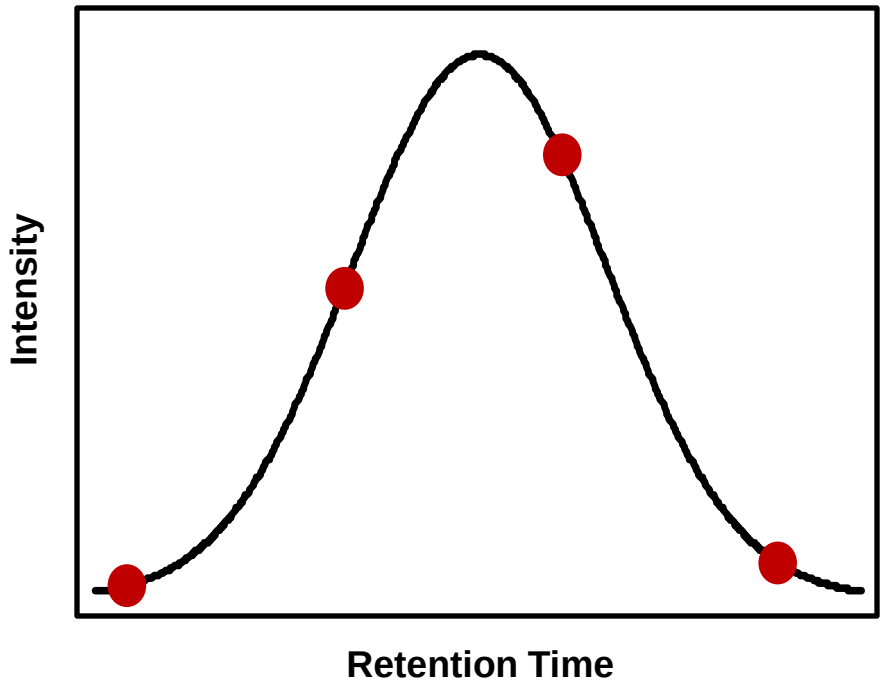


Time dimension

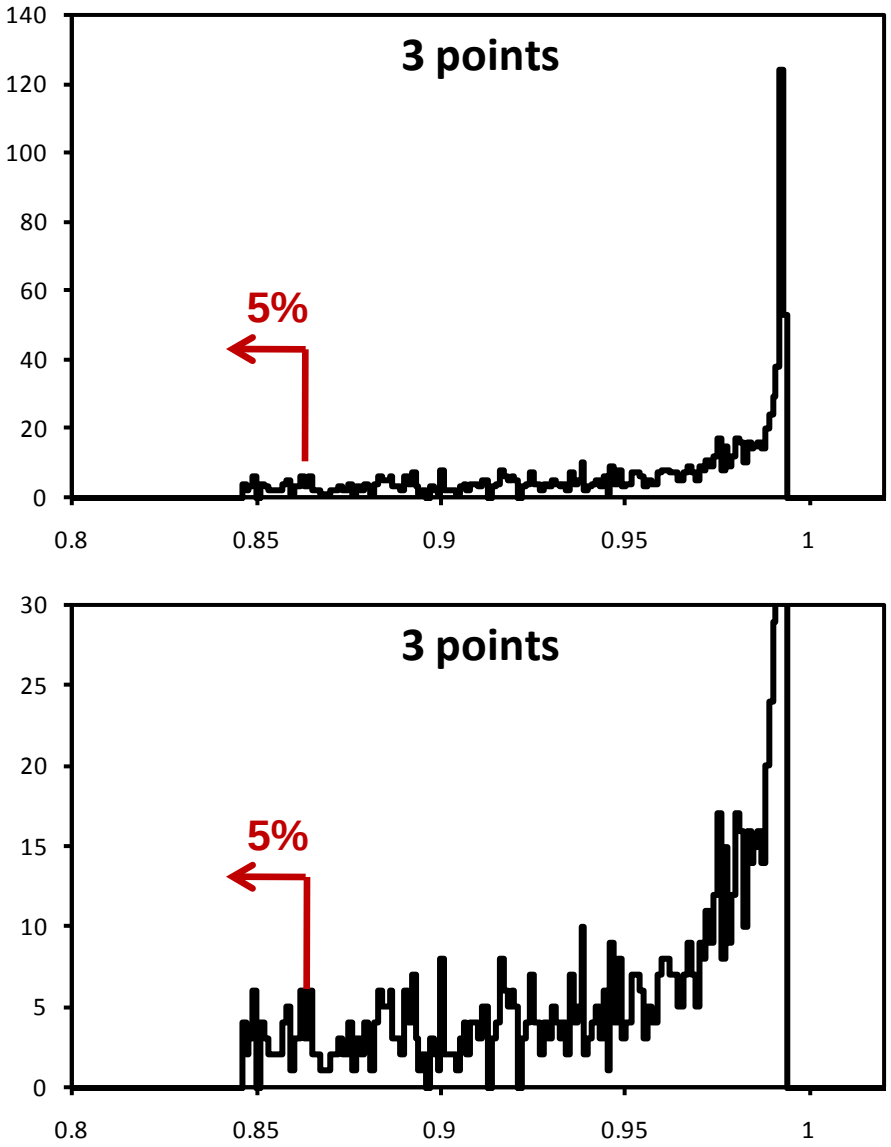




Sampling

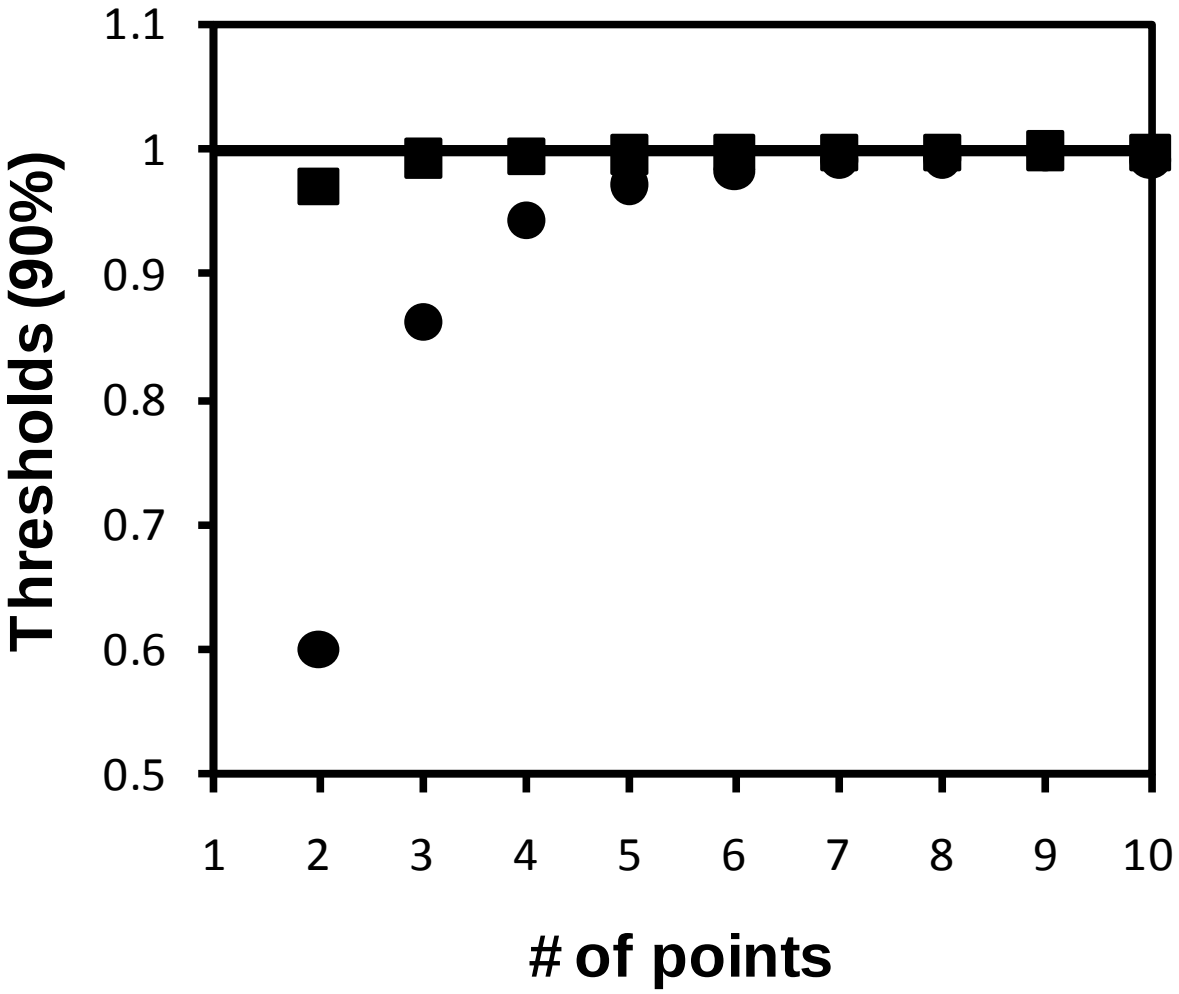


Sampling



Acquisition time = 0.05σ

Sampling

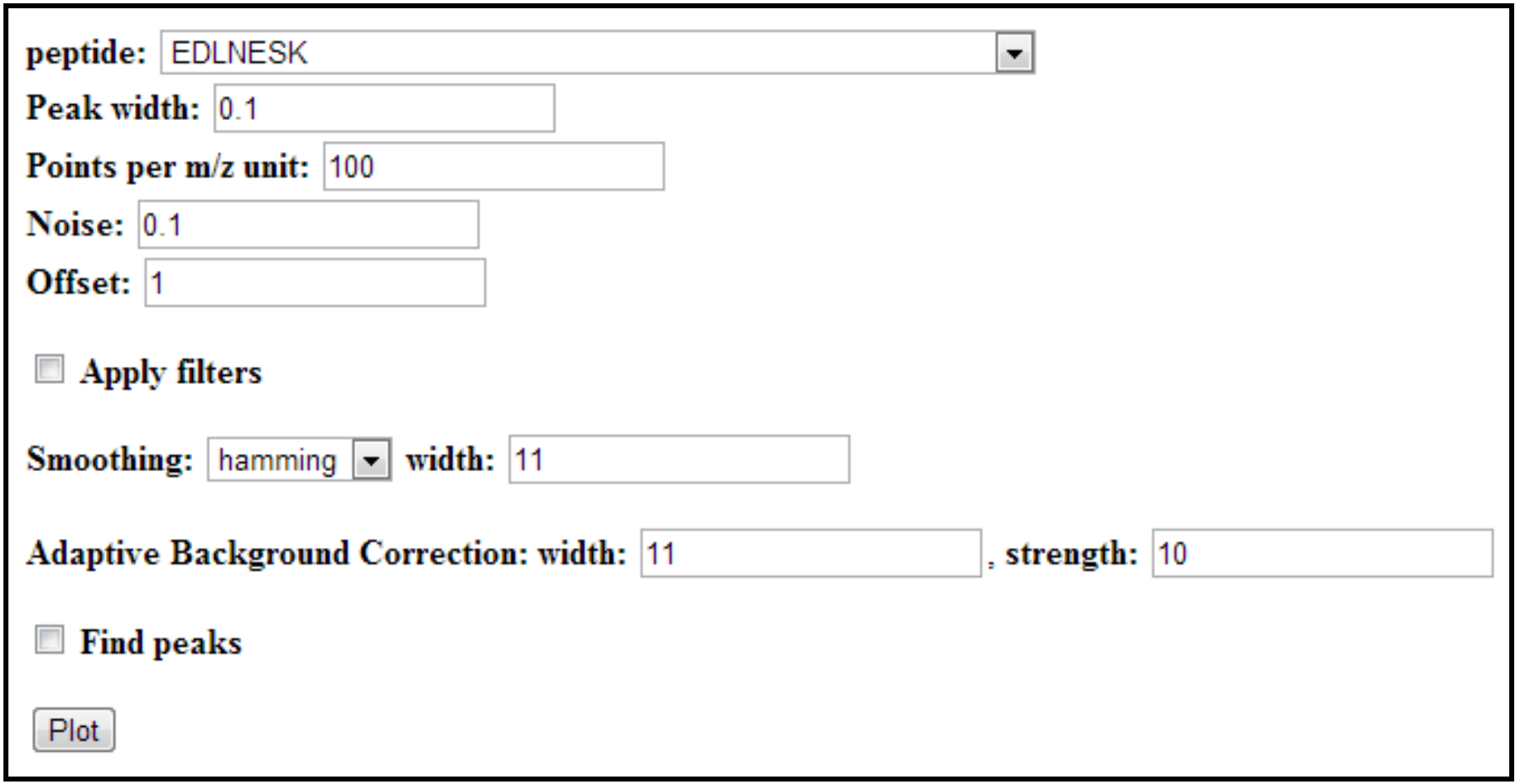


What is the best way to estimate quantity?

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

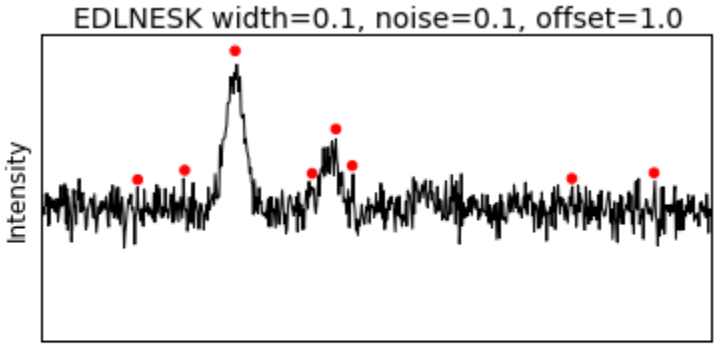
Web Tool

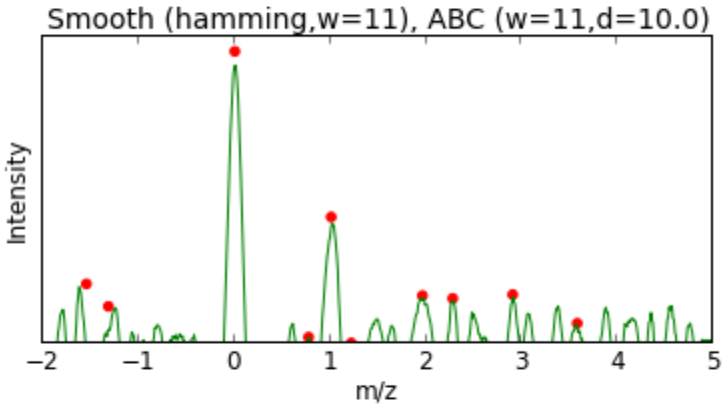
http://10.193.36.101/plot-filter-cgi/plot_filter.pl or
http://10.193.36.219/plot-filter-cgi/plot_filter.pl



Web Tool

http://10.193.36.101/plot-filter-cgi/plot_filter.pl or
http://10.193.36.219/plot-filter-cgi/plot_filter.pl





peptide mass = 833.37668
m/z = 834.38395 for z=1
m/z = 417.69561 for z=2
m/z = 278.7995 for z=3
m/z = 209.35145 for z=4

Peaks found (original):
m/z=-1.0178, -0.52549, 0.0011996, 0.80683, 1.0546, 1.2254, 3.5219, 4.3809,
Peaks found (filtered):
m/z=-1.5527, -1.3224, -0.0013772, 0.76976, 1.0058, 1.2151, 1.9598, 2.2765, 2.9036, 3.5779,

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