MS-DIAL FAQ

Hiroshi Tsugawa

How it works for peak detections

- ✓Differential function
- ✓Noise estimation

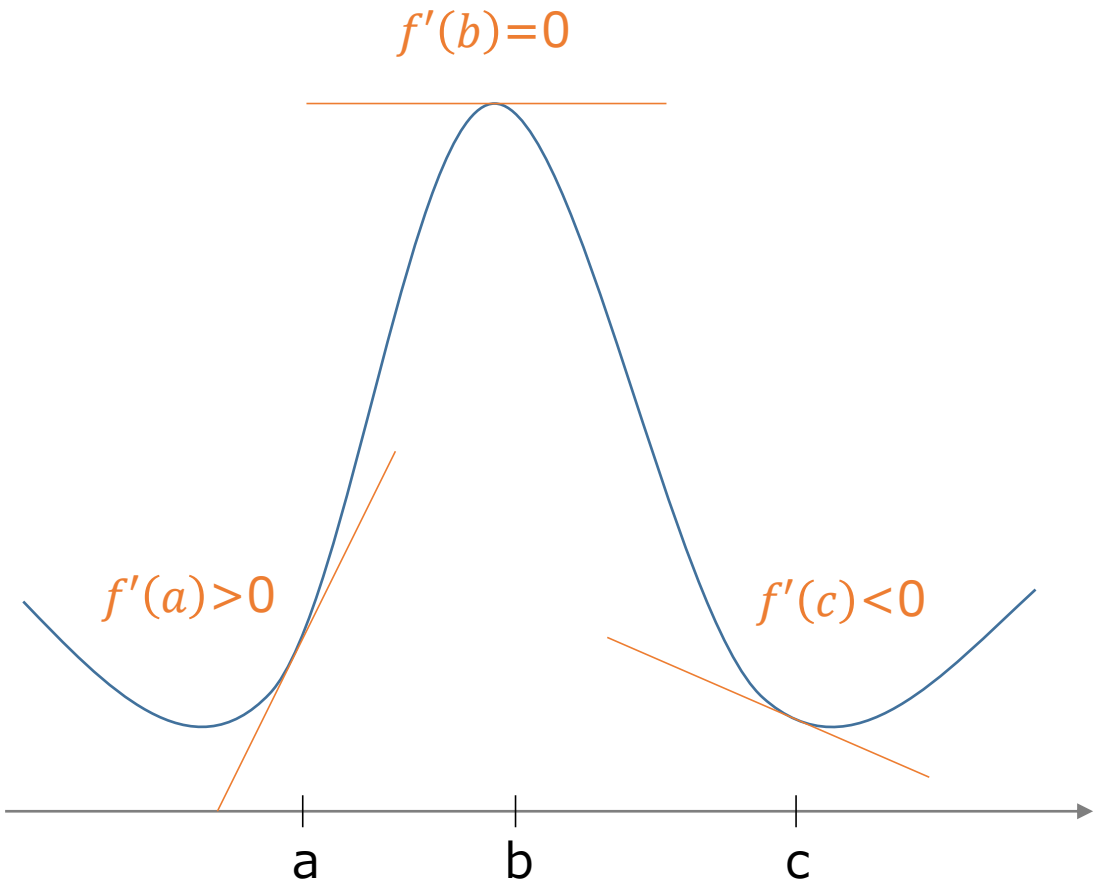
Differential function

$$f(x) = ax^3 + bx^2 + cx + d$$

$$\downarrow$$

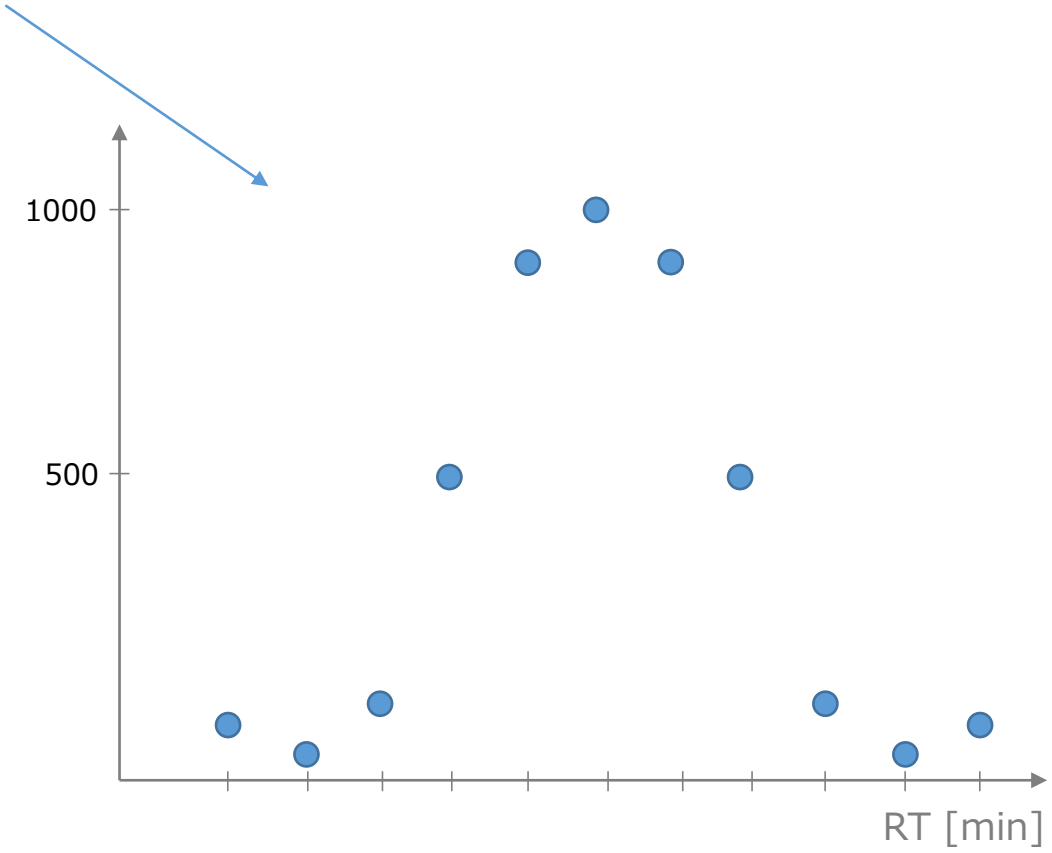
$$f'(x) = 3ax^2 + 2bx + c$$

Slope



Differential function for chromatogram

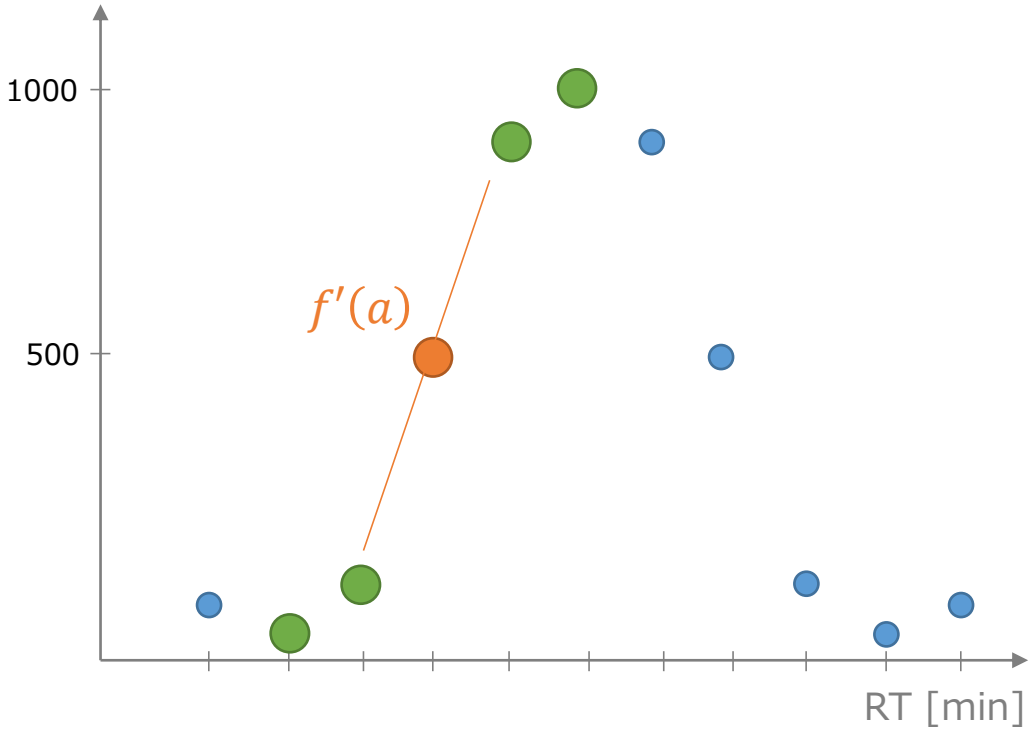
$f(x) = \{50, 10, 100, 500, 800, 1000, 800, 500, 100, 10, 50\}$



Differential function for chromatogram

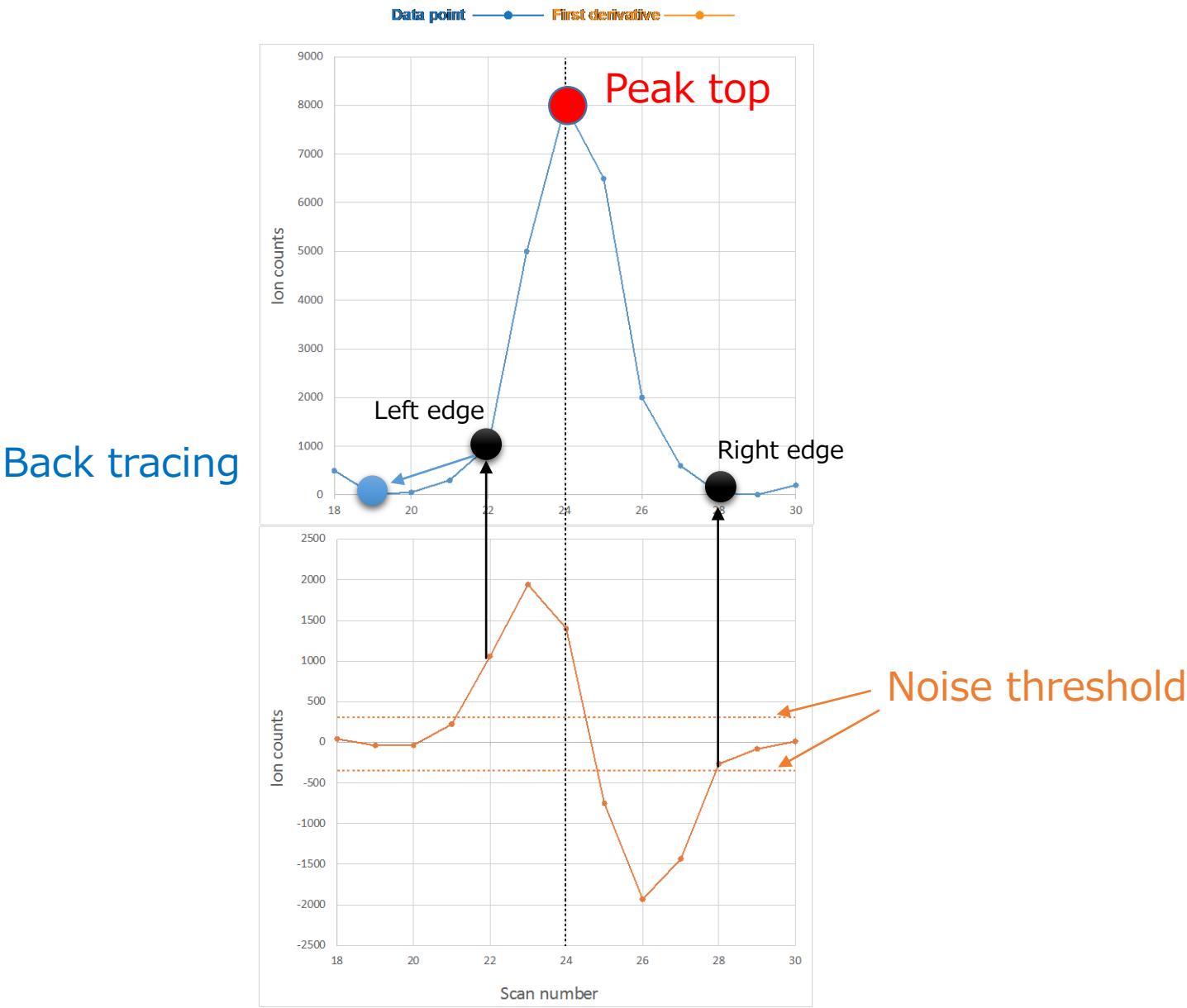
$$f(x) = \{50, 10, 100, 500, 800, 1000, 800, 500, 100, 10, 50\}$$

$$f'(x) = \frac{-2x_{-2} - x_{-1} + x_{+1} + 2x_{+2}}{10}$$

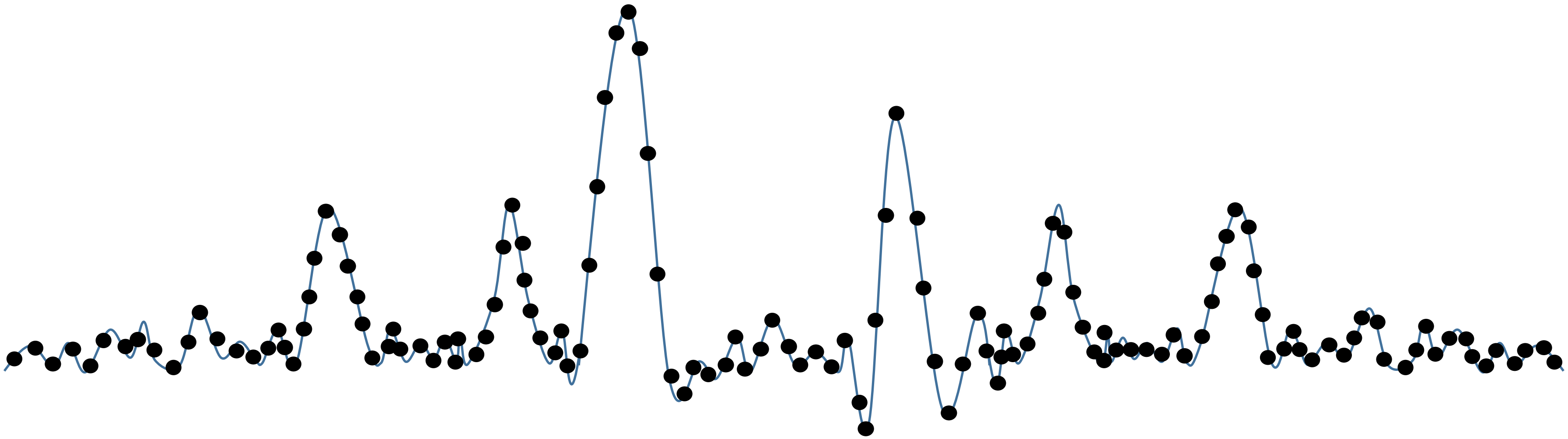


Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627-1639 (1964).

Differential function for chromatogram



Noise evaluation



$$f'(x) = \{10,20,5,-5,-1,10,100,1000,3000,\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..,5,-10\}$$

Sort the first derivative array by the absolute values

Noise evaluation

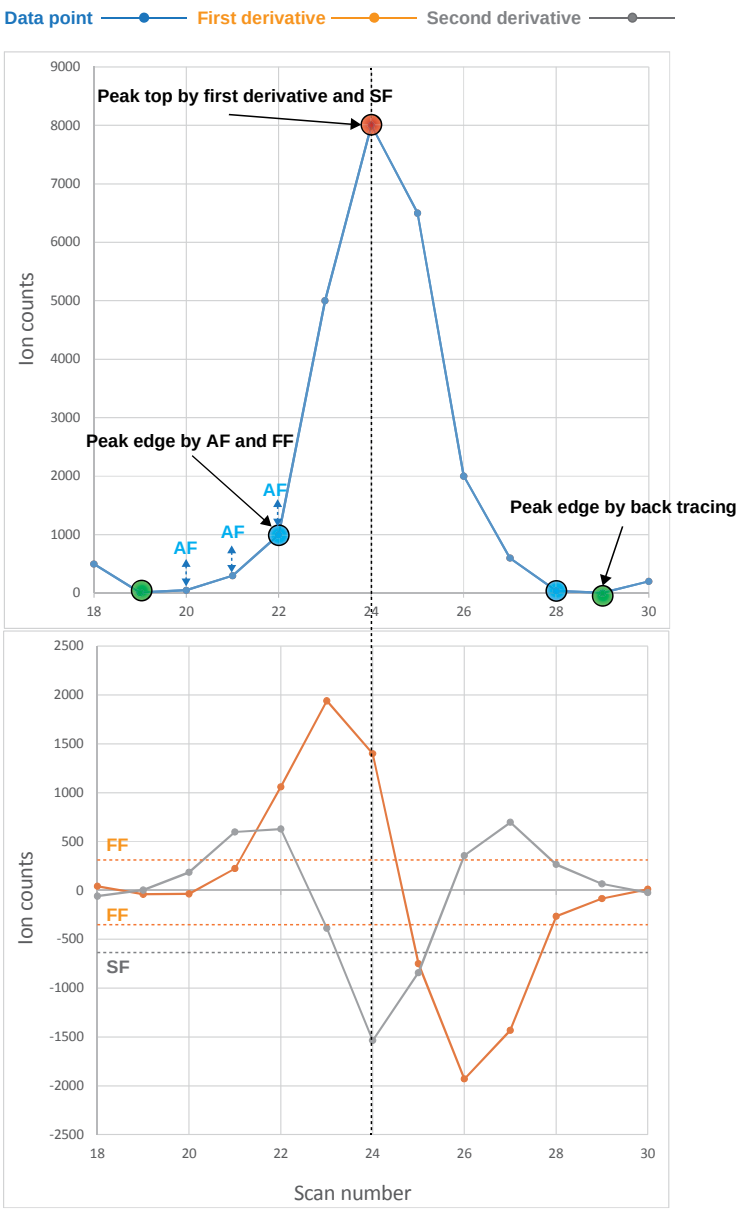
Noise value = Median

f'(x)
0.361563
2.368109
6.641594
9.488878
10.547
11.025
16.82708
17.22223
21.78178
22.57424
23.91914
27.35502
30.6929
31.75519
32.60672
36.23518
40.57042
45.80185
47.27034
55.0006
58.68845
58.71436
61.12281
61.28744
70.8877
73.31488
77.39816
77.75772
77.99117
80.14925
85.77214
86.85782
⋮

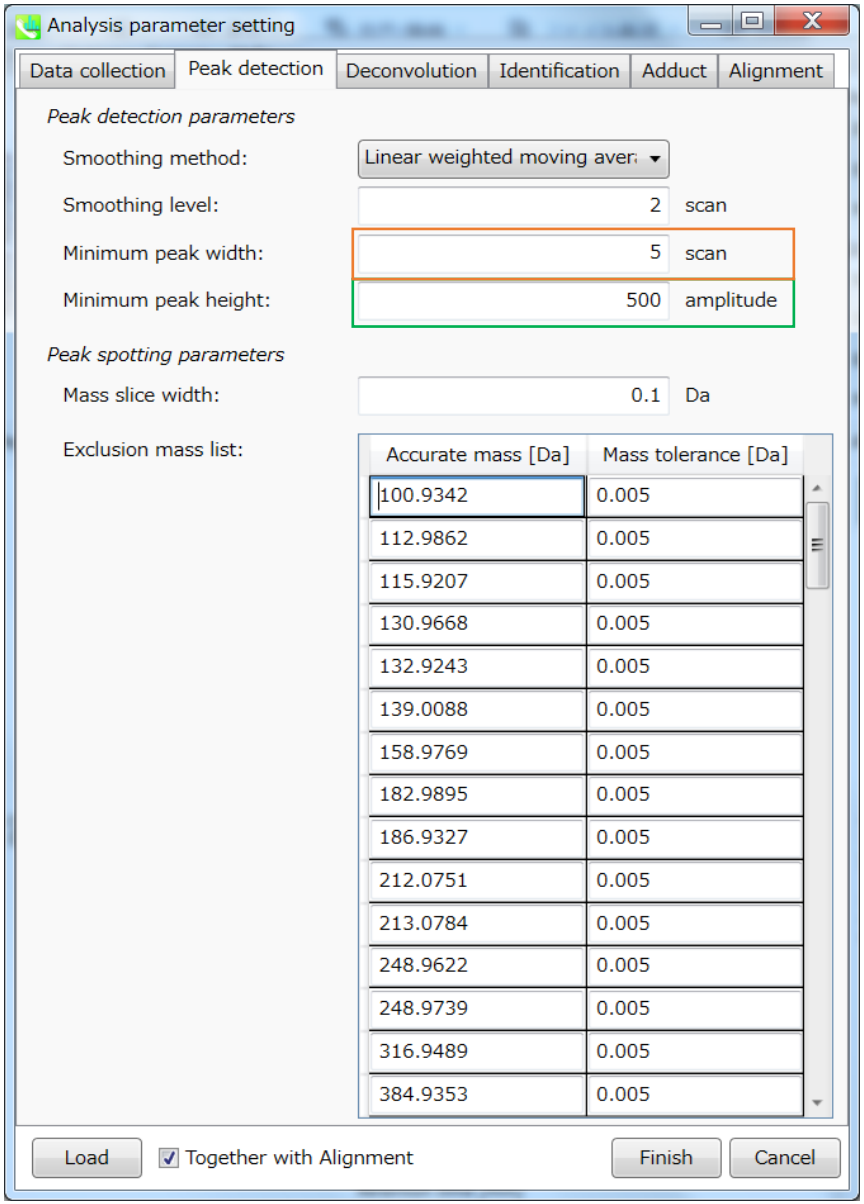
Top 5%

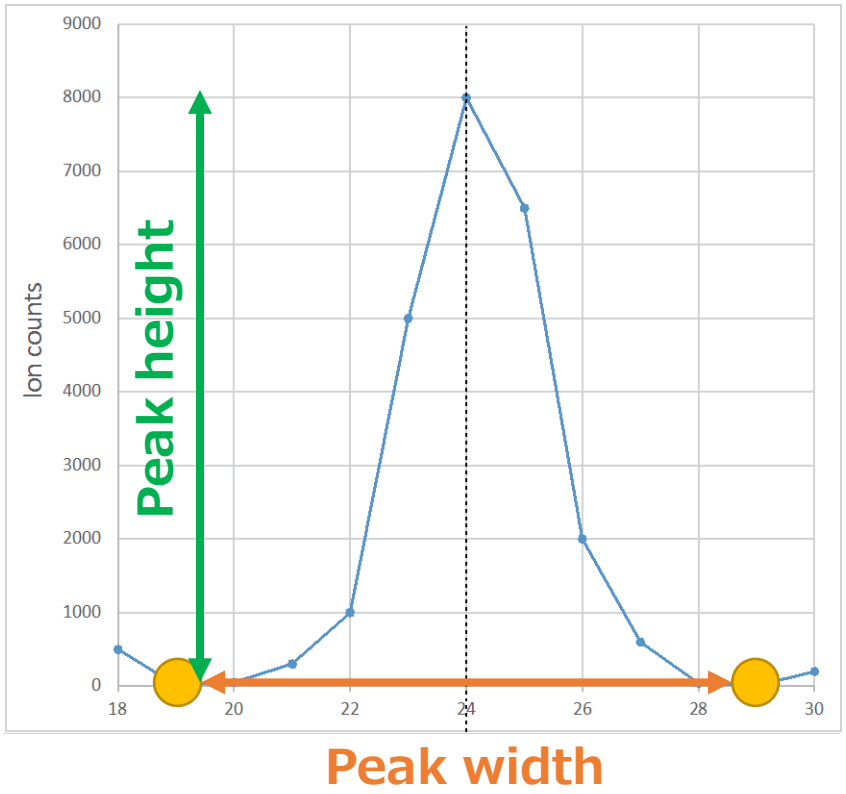
In MS-DIAL program

Not only first derivatives,
But also the amplitude differences
and the second derivatives are evaluated.

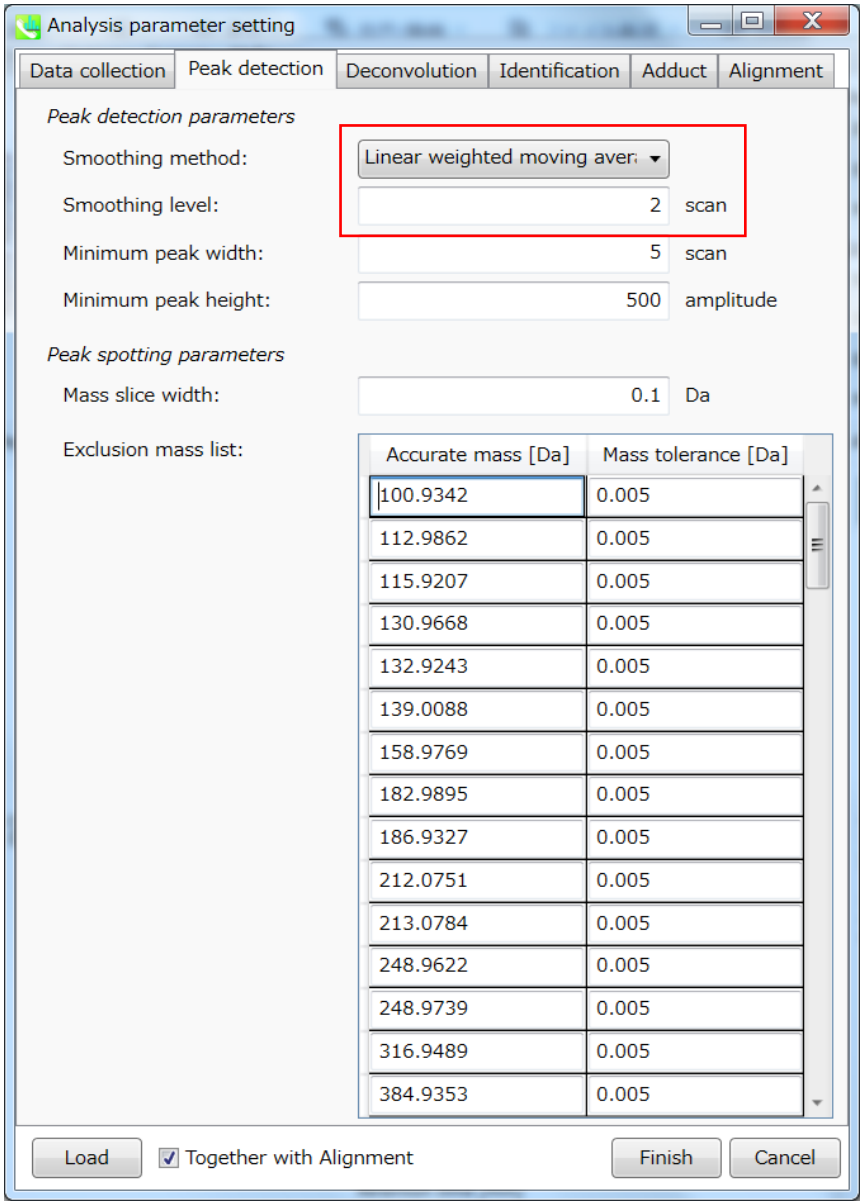


In MS-DIAL program

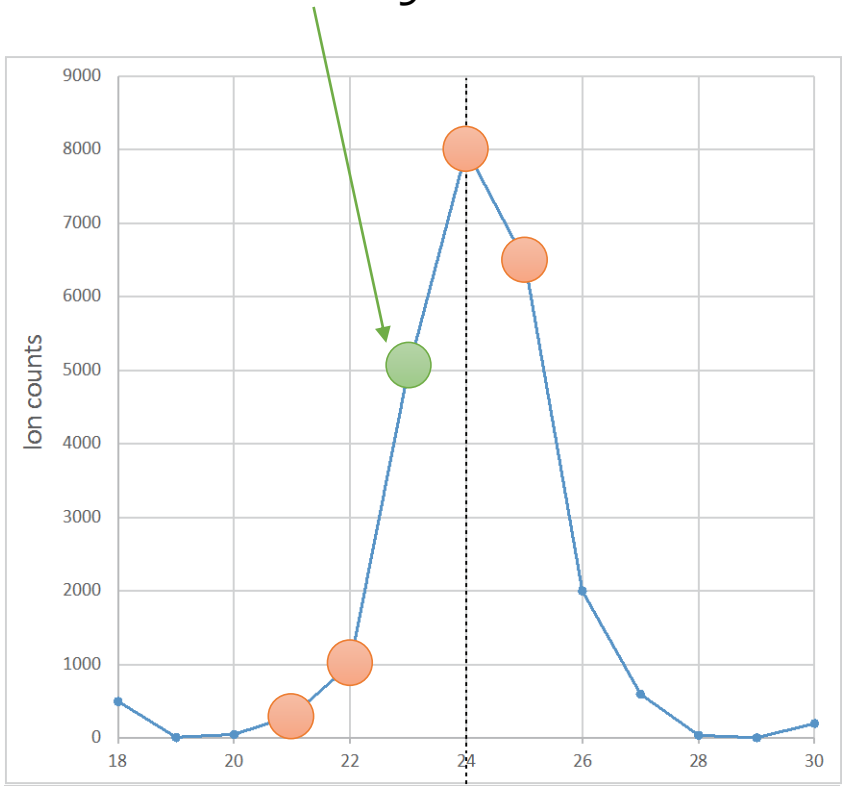




In MS-DIAL program



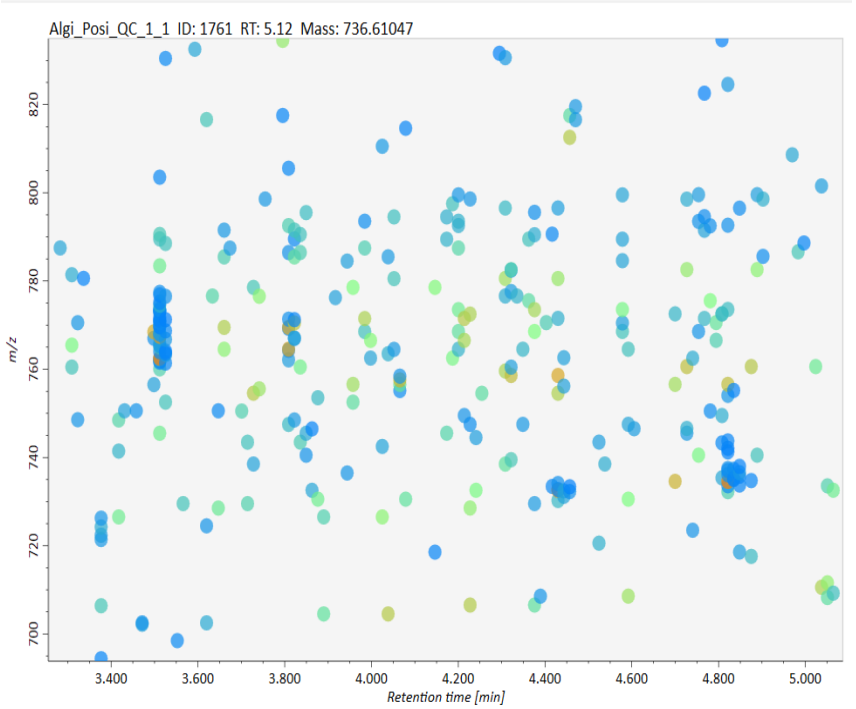
$$smoothed\ value = \frac{1f(x_{-2}) + 2f(x_{-1}) + 3f(x) + 2f(x_{+1}) + 1f(x_{+2})}{9}$$



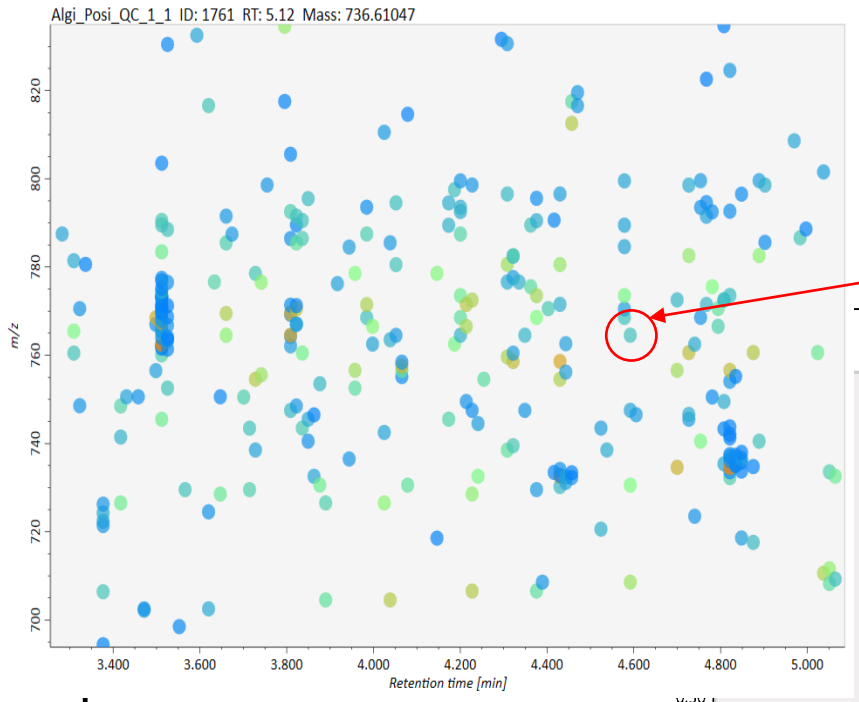
How it works in RT & *m/z* axis

✓**Slicing method**

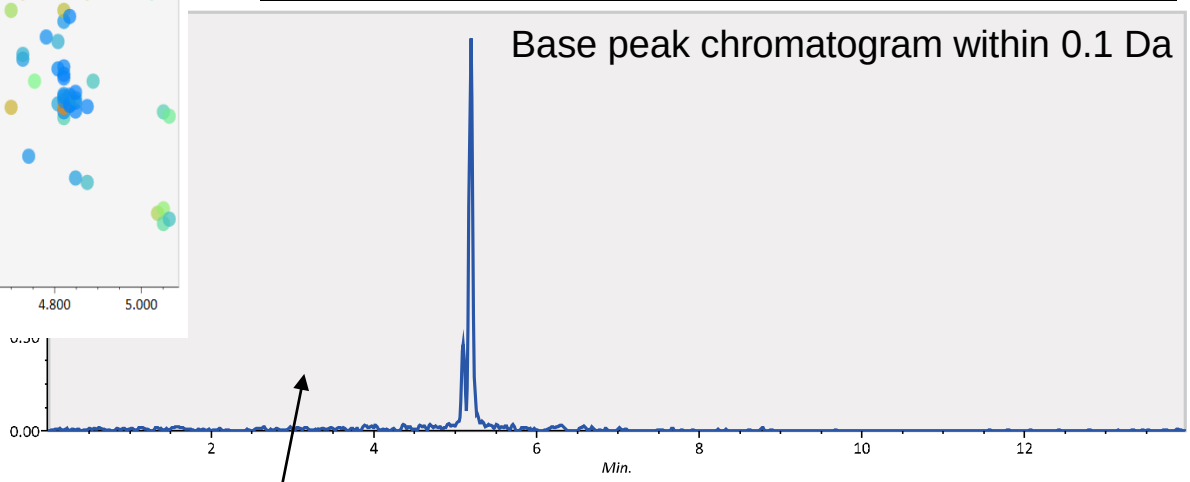
✓**Peak detection is applied to**
base peak chromatogram

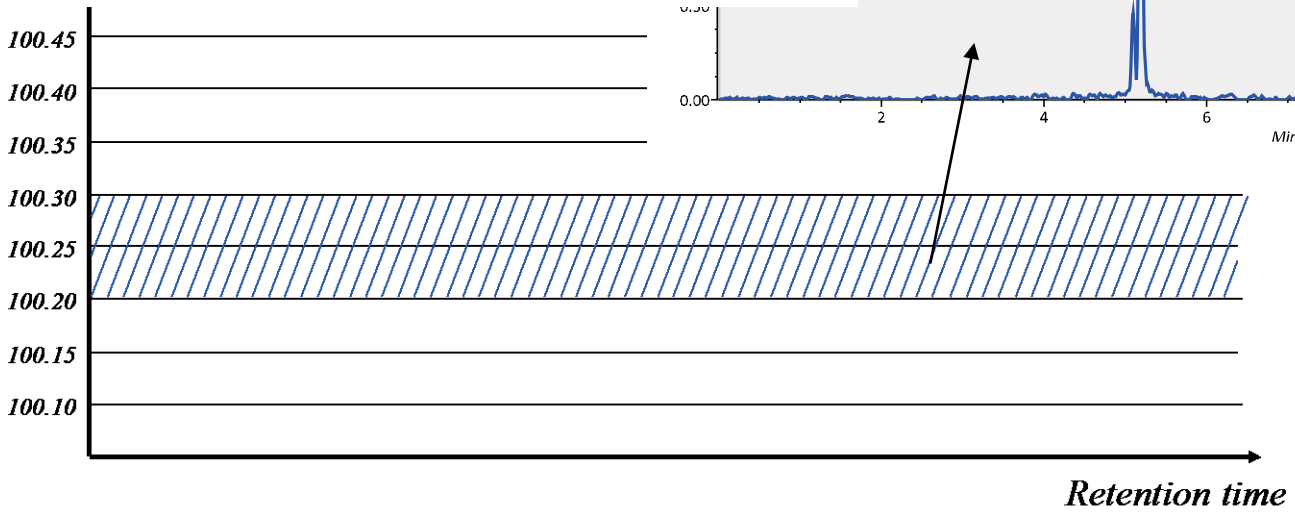


Slicing method & BPC exctraction

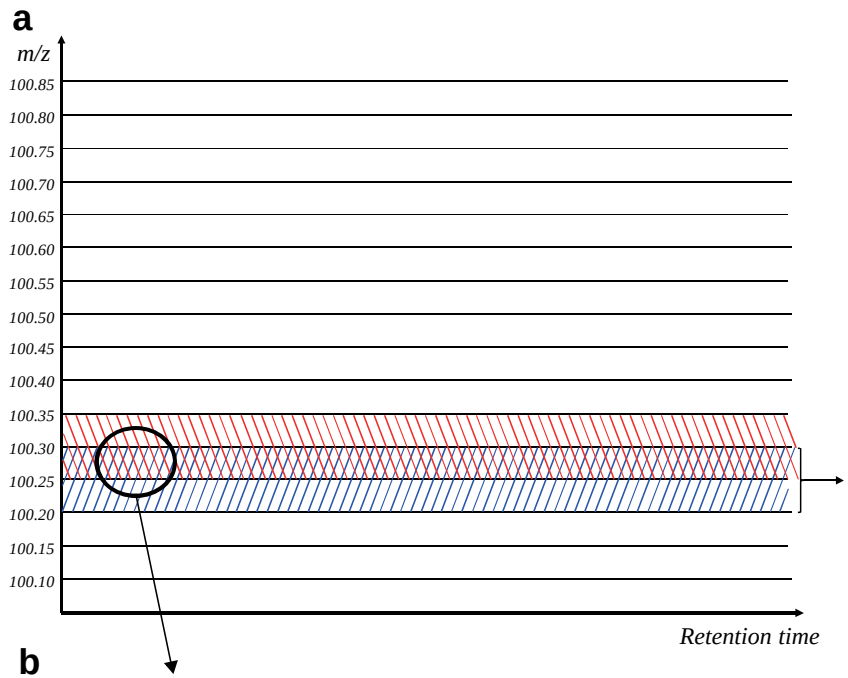


Scan number	Retention time [min]	Base peak *m/z*	Base peak intensity
1	0.1	100.2054	1
2	0.12	100.2053	10
3	0.14	100.2053	5
4	0.16	100.2052	50
5	0.18	100.2051	200
6	0.2	100.2054	1500
7	0.22	100.2054	3000
8	0.24	100.2054	1700
9	0.26	100.2053	180
10	0.28	100.205	60

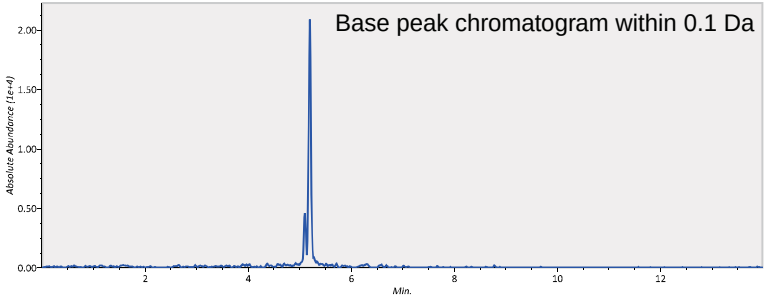


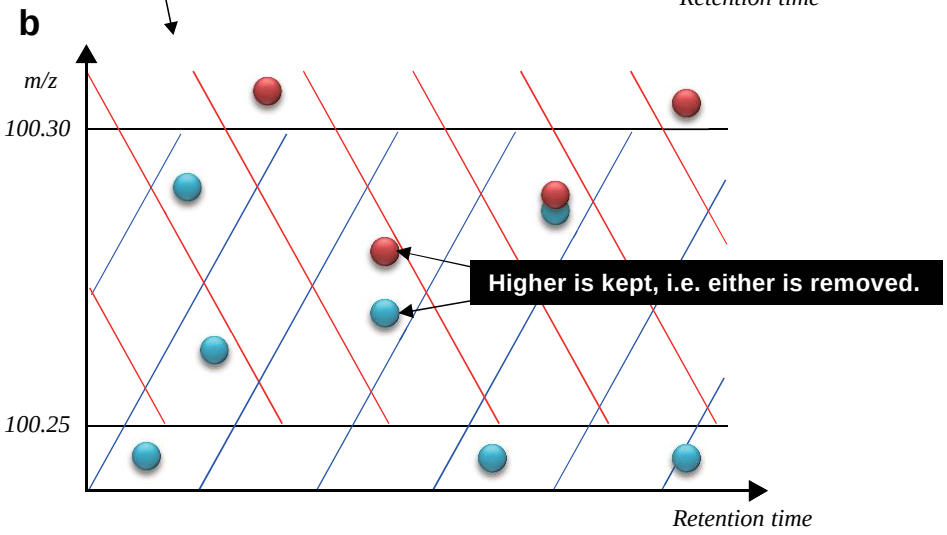


Slicing method & BPC exctraction

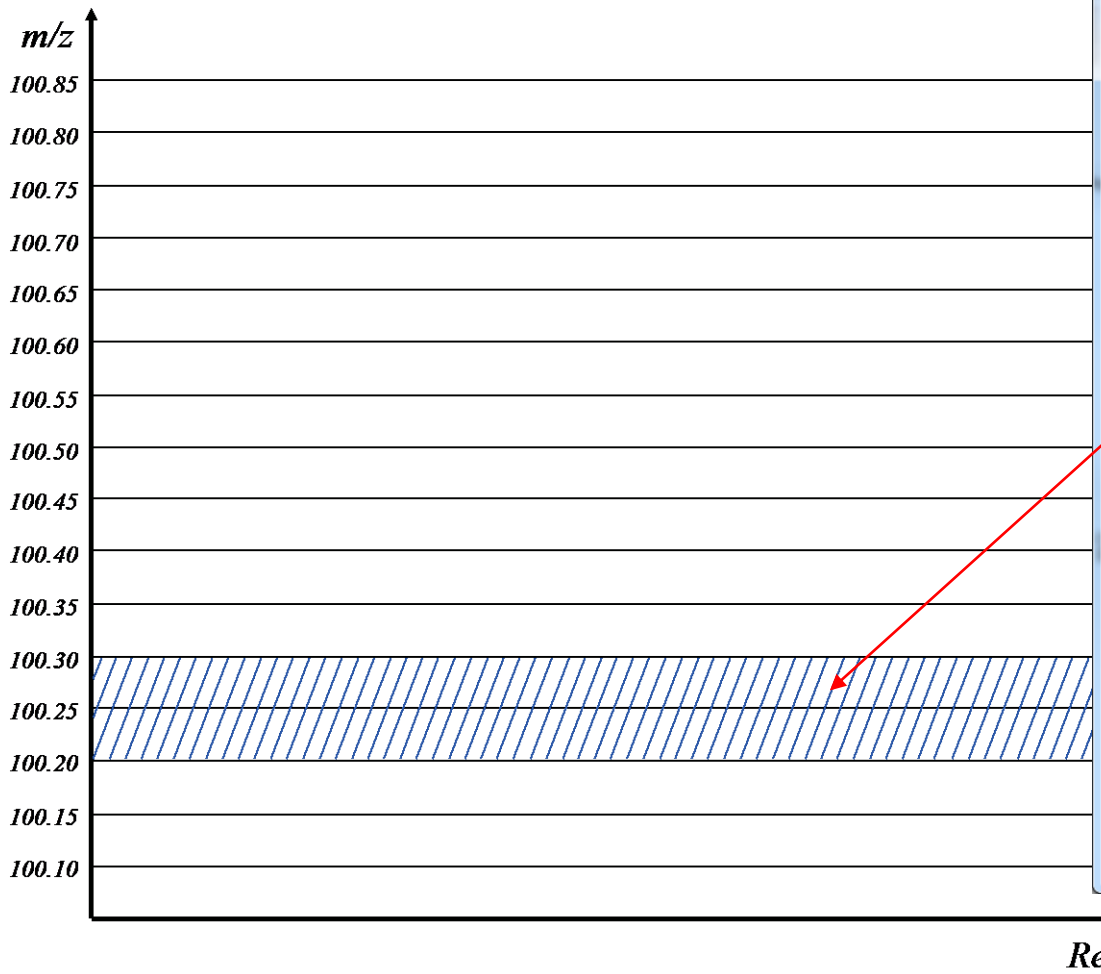


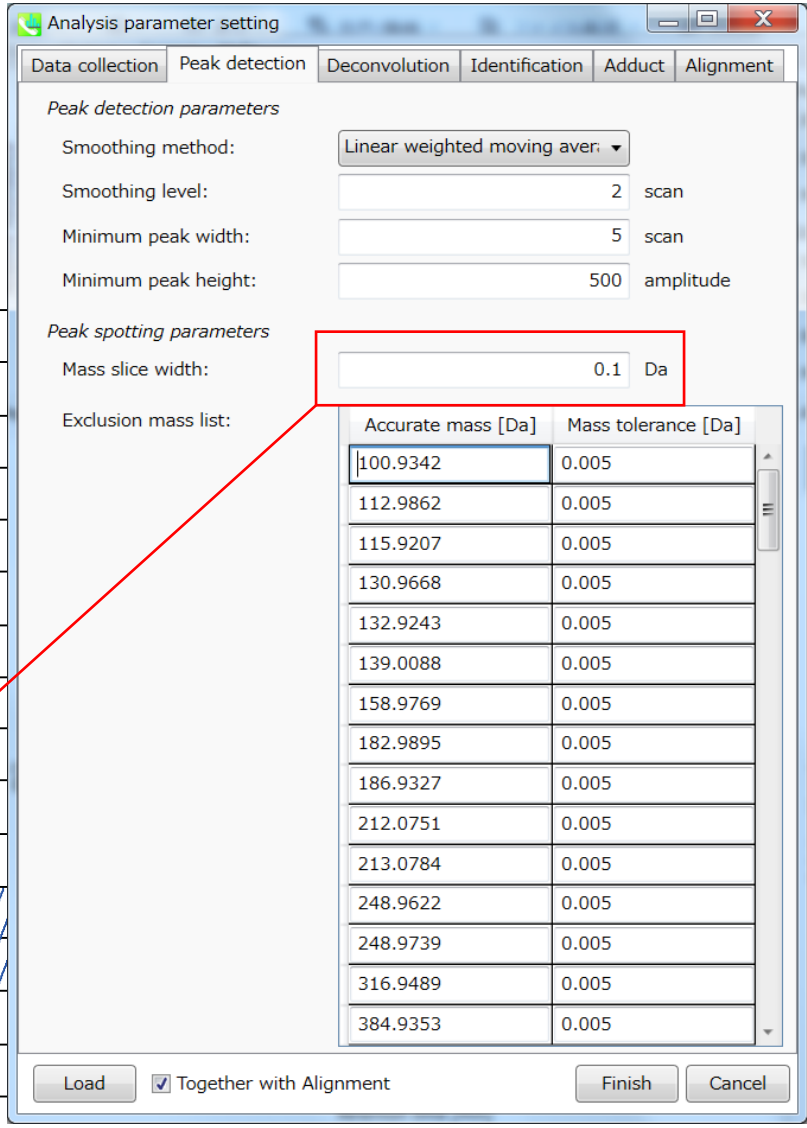
Scan number	Retention time [min]	Base peak *m/z*	Base peak intensity
1	0.1	100.2054	1
2	0.12	100.2053	10
3	0.14	100.2053	5
4	0.16	100.2052	50
5	0.18	100.2051	200
6	0.2	100.2054	1500
7	0.22	100.2054	3000
8	0.24	100.2054	1700
9	0.26	100.2053	180
10	0.28	100.205	60



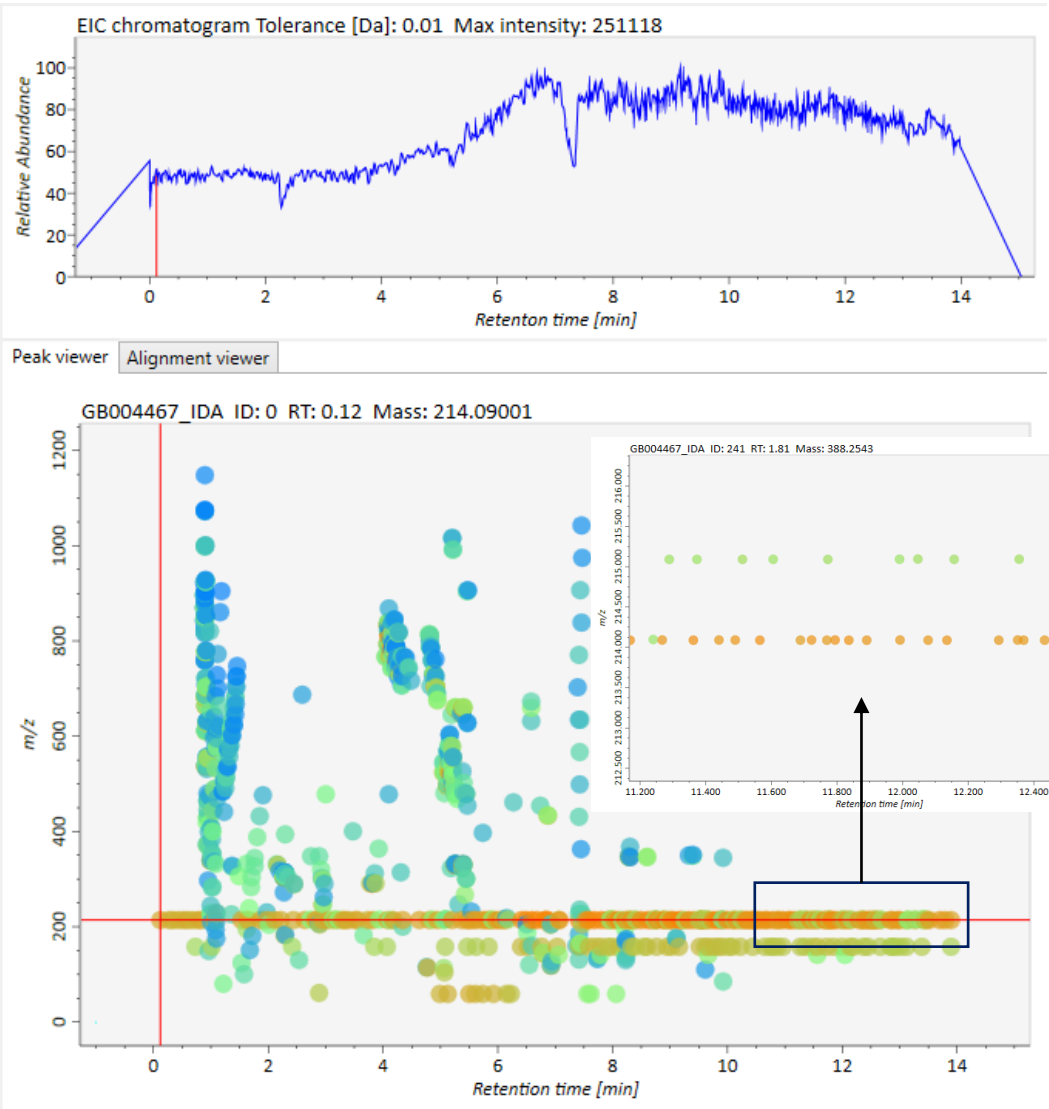


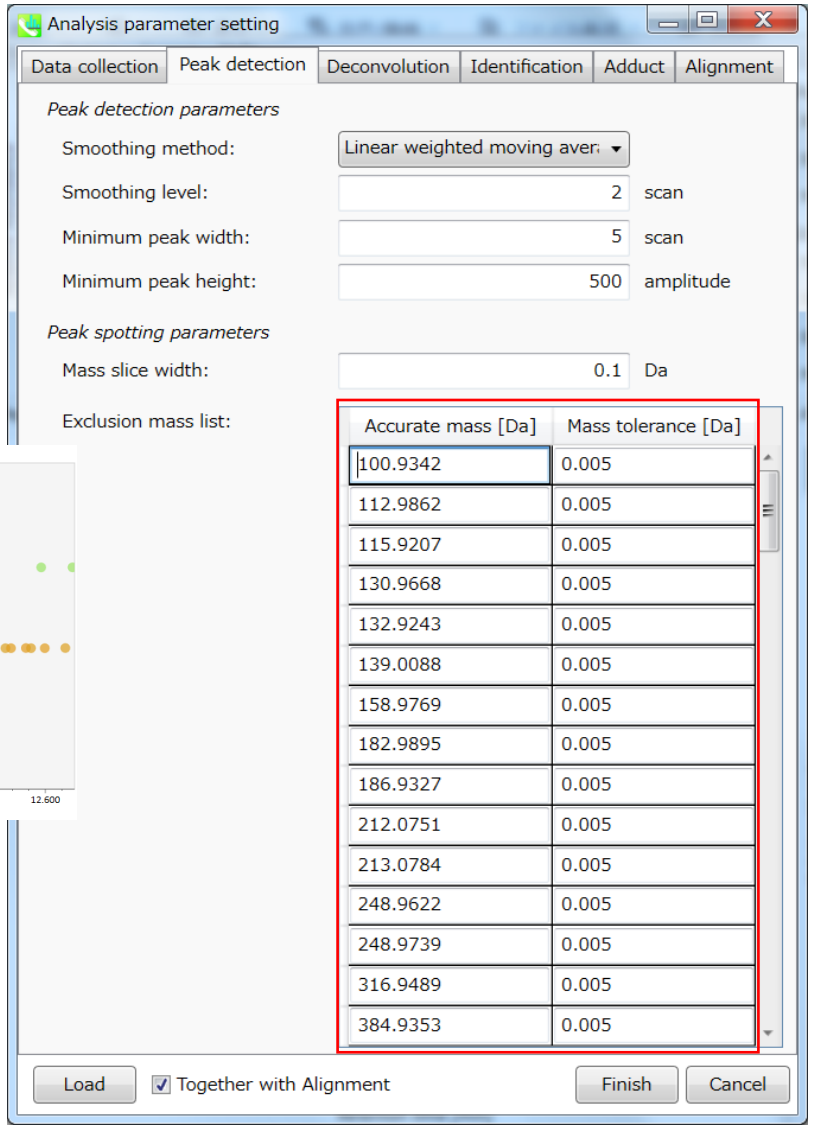
Slicing method & BPC exctraction



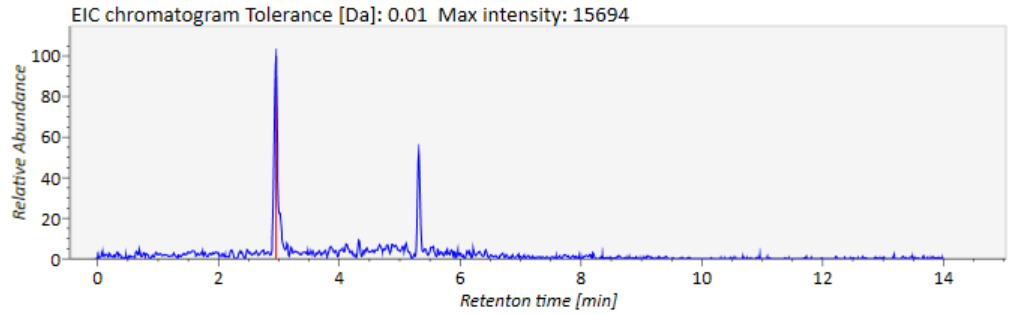


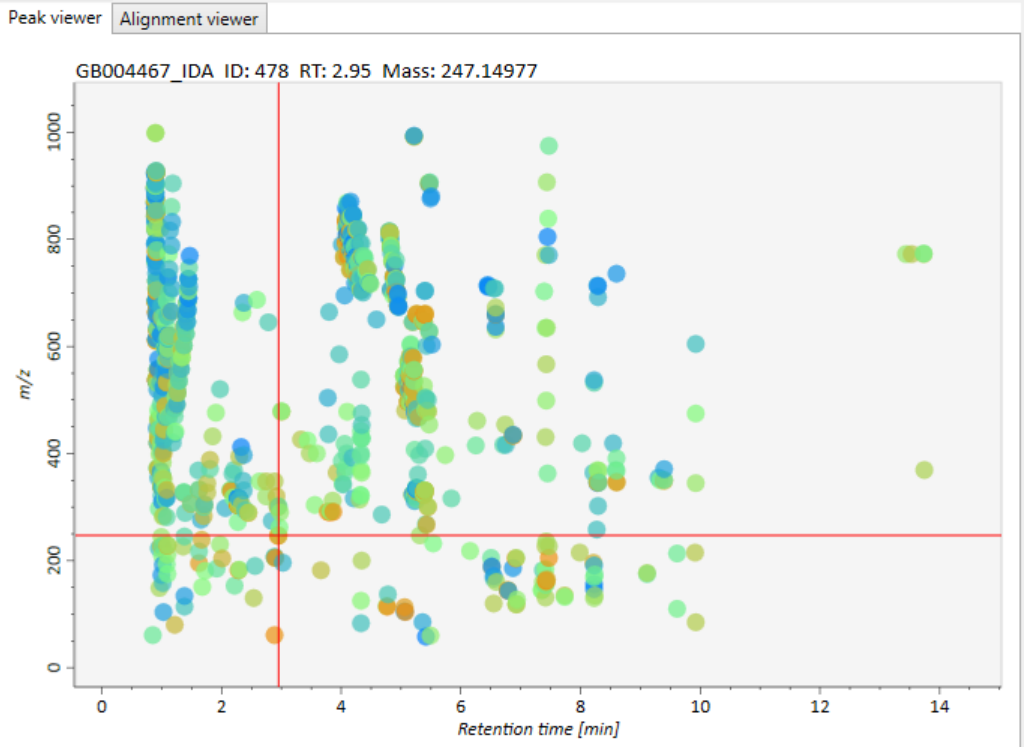
What is ‘exclusion mass list’?

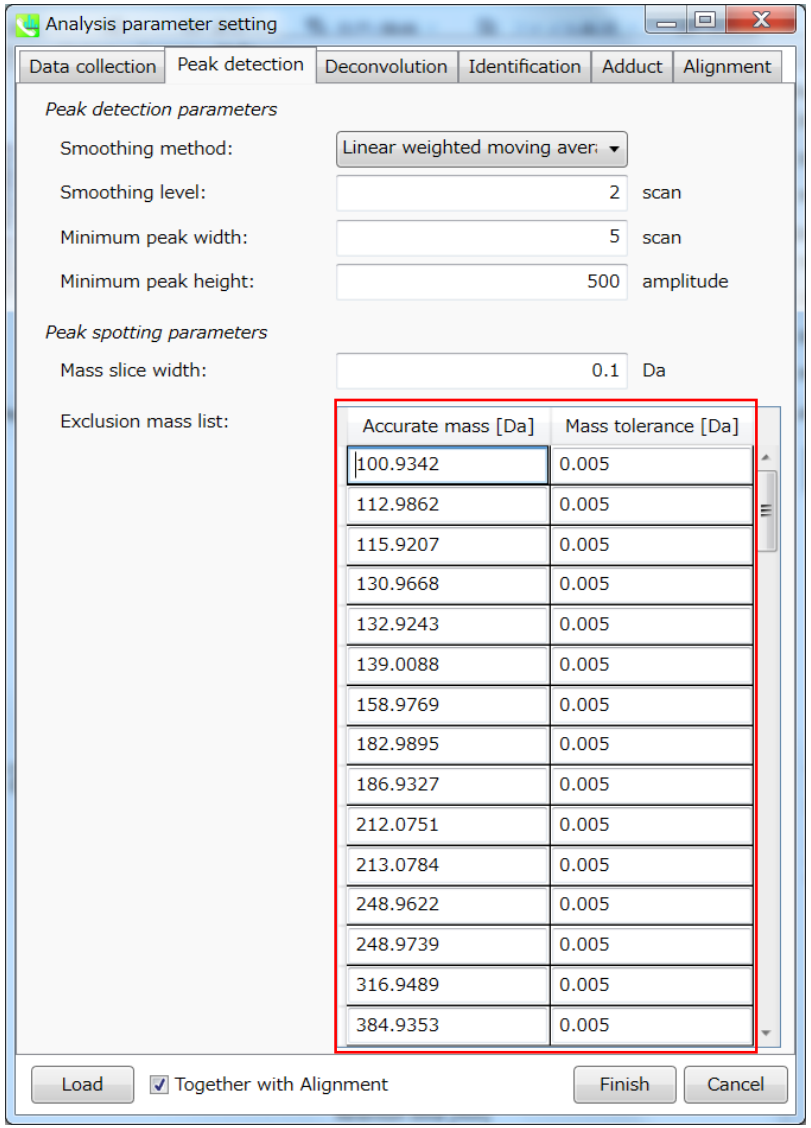




What is ‘exclusion mass list’?

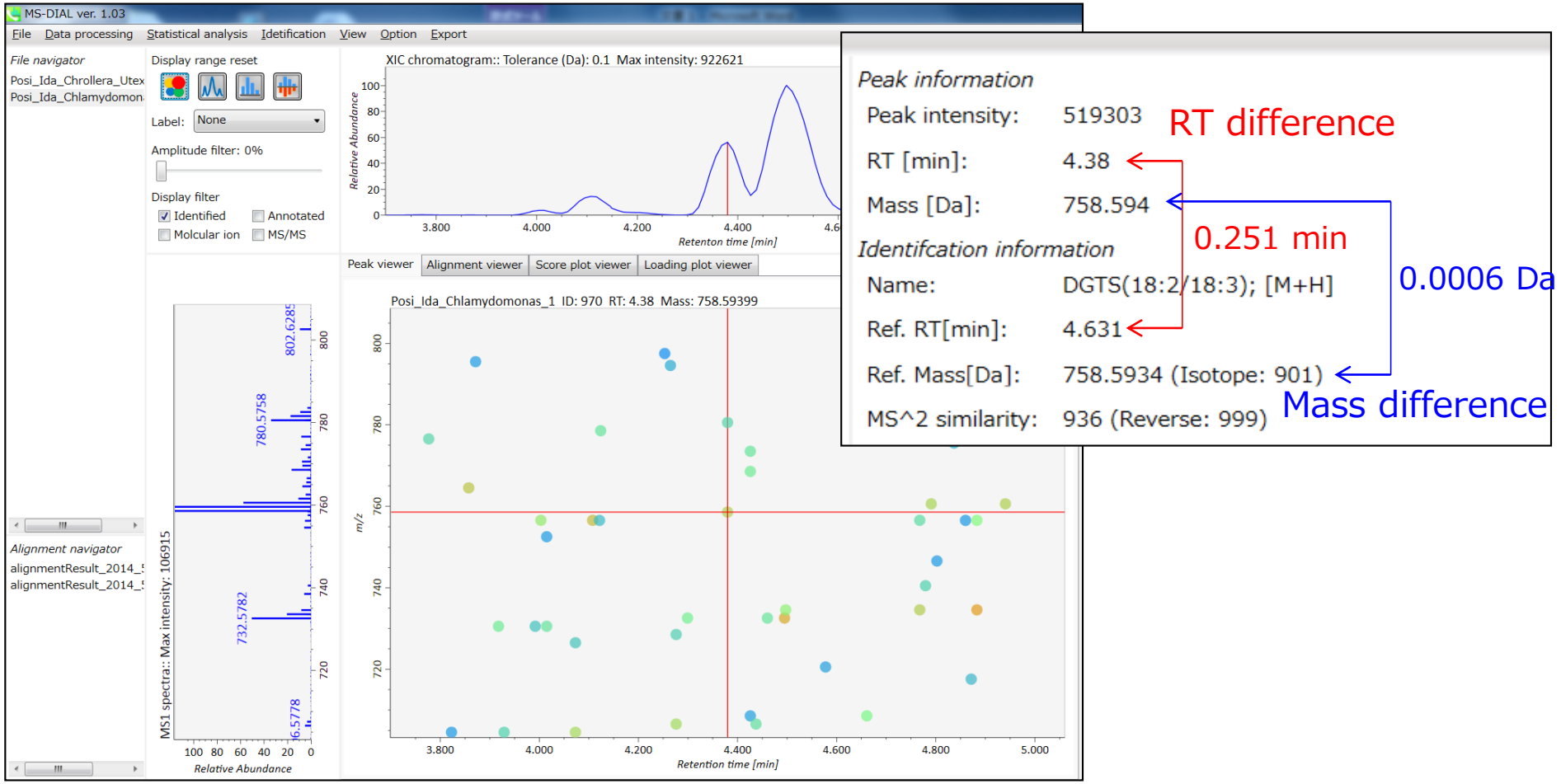






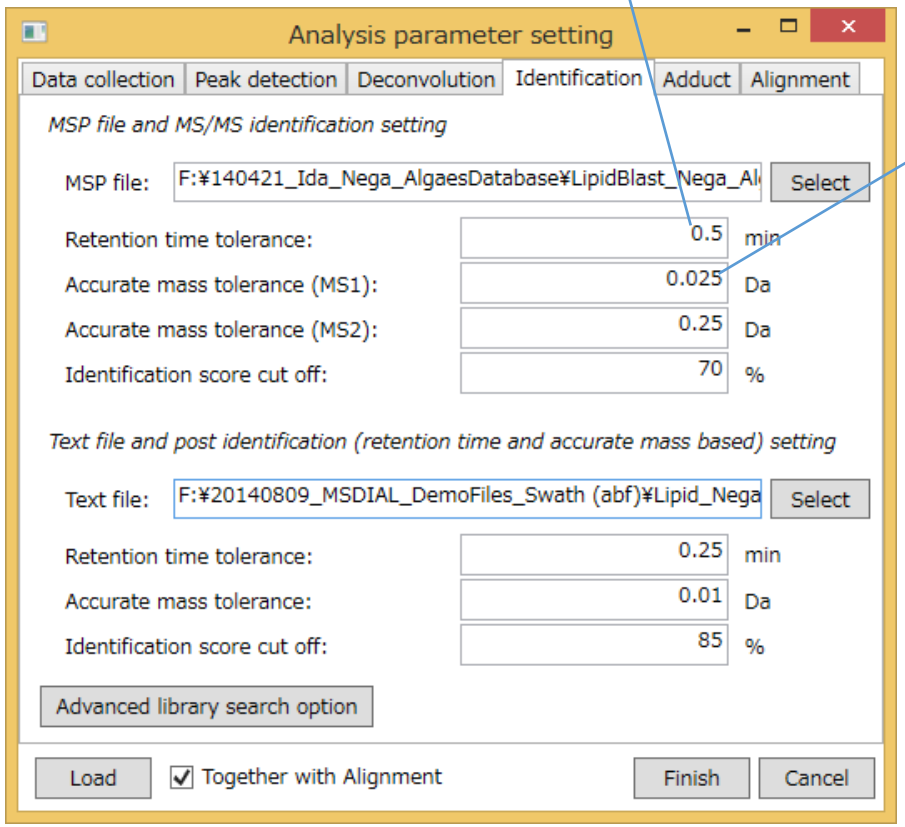
Retention time and MS1 similarity

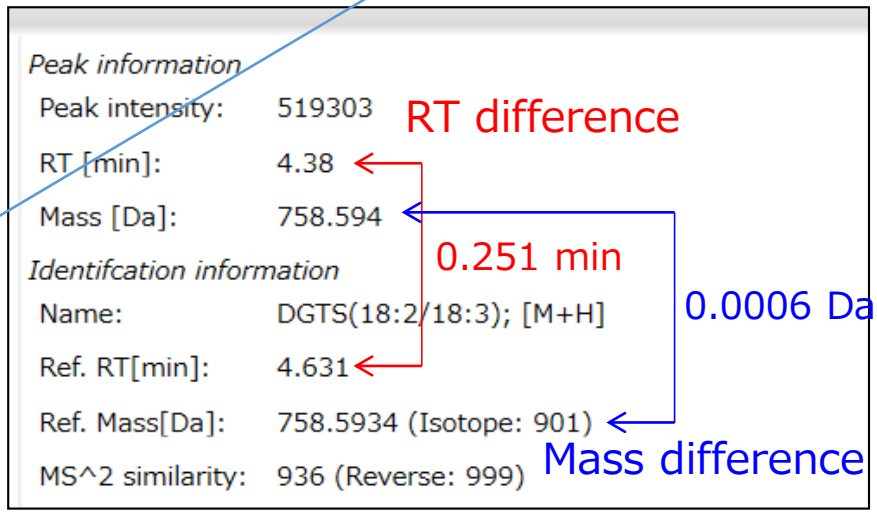
$$S_{RT} = \exp\left\{-0.5 \times \left(\frac{RT_{act.} - RT_{lib.}}{\delta}\right)^2\right\} \qquad S_{MS1} = \exp\left\{-0.5 \times \left(\frac{Mass_{act.} - Mass_{lib.}}{\delta}\right)^2\right\}$$



Retention time and MS1 similarity

$$S_{RT} = \exp\left\{-0.5 \times \left(\frac{RT_{act.} - RT_{lib.}}{\delta}\right)^2\right\} \qquad S_{MS1} = \exp\left\{-0.5 \times \left(\frac{Mass_{act.} - Mass_{lib.}}{\delta}\right)^2\right\}$$



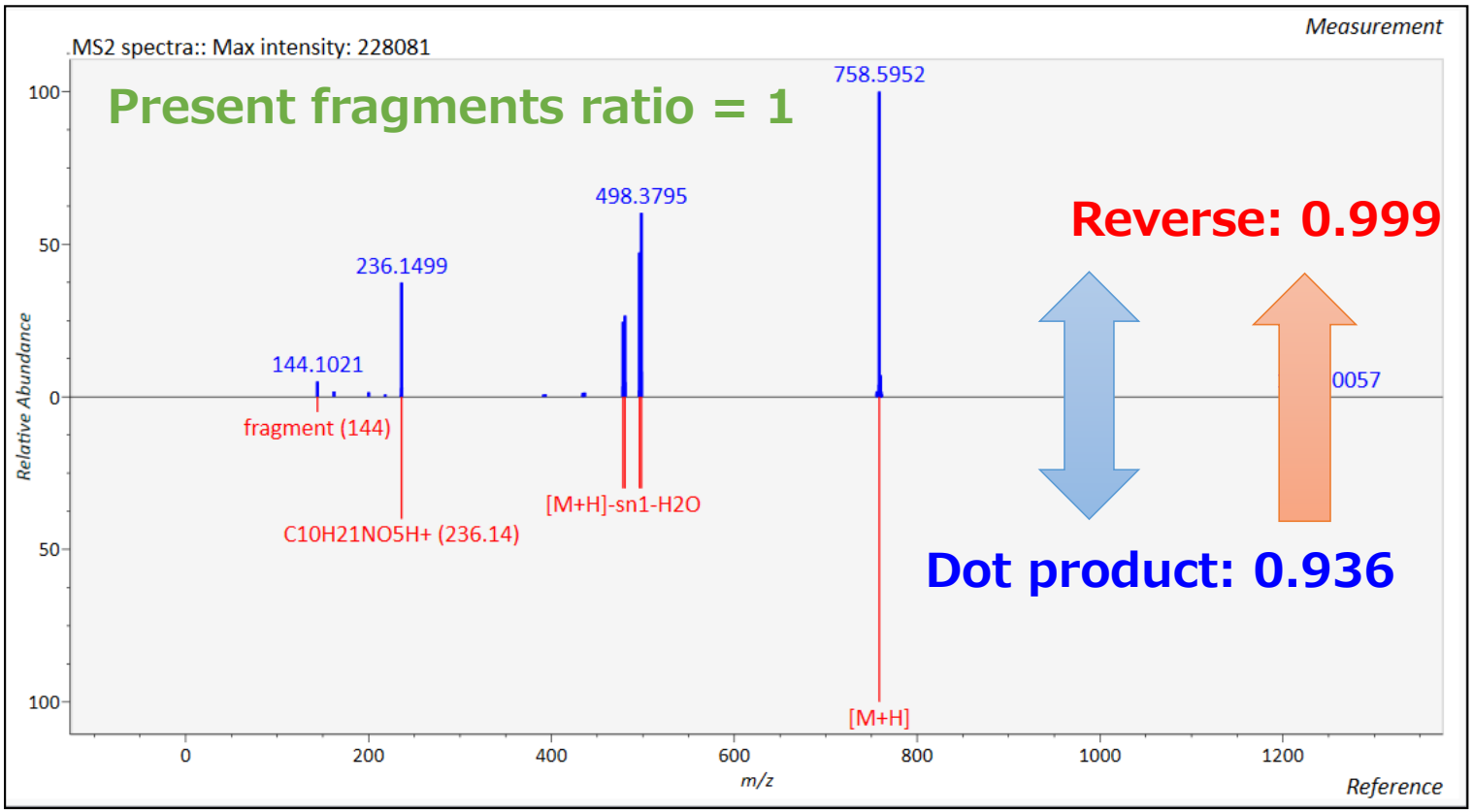


MS/MS similarity

$$dot\ product = \frac{(\sum W_{act.} W_{lib.})^2}{\sum {W_{act.}}^2 \sum {W_{lib.}}^2} \qquad dot\ product_{reverse} = \frac{(\sum W_{act.} W_{lib.})^2}{\sum {W_{act.}}^2 \sum {W_{lib.}}^2} \ (\text{"in lib."only})$$

Amplitude(A) is normalized by $1/\left(1+\frac{A}{\sum A-0.5}\right)$

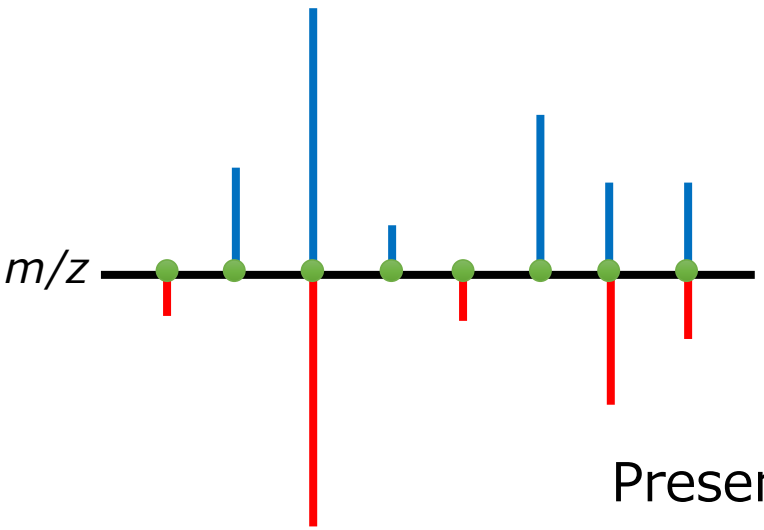
$$Present\ fragments\ ratio = \frac{The\ number\ of\ mathed\ fragments}{The\ number\ of\ reference\ fragments}$$



BTW, what is the difference?

Dot product

Act.: **a** = (0, 400, 1000, 200, 0, 700, 300, 300)

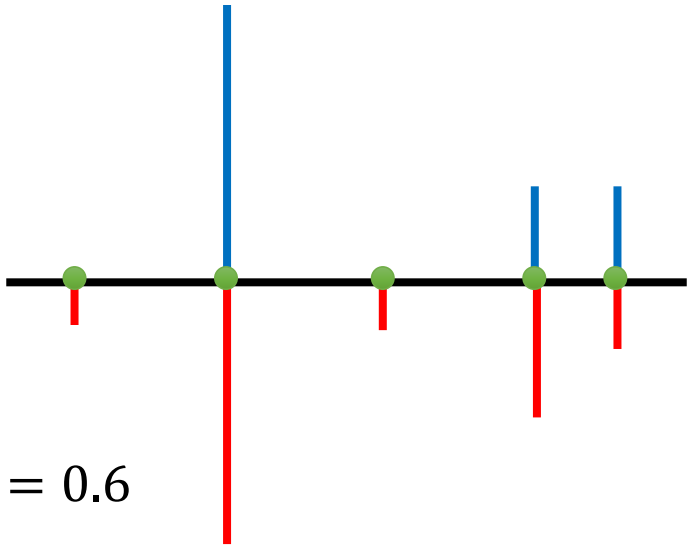


Lib: **b** = (100, 0, 1000, 0, 100, 0, 400, 300)

Dot product: $\frac{a \cdot b}{|a| \cdot |b|} = 0.637$

Reverse dot product

Act.: **a'** = (0, 1000, 0, 300, 300)



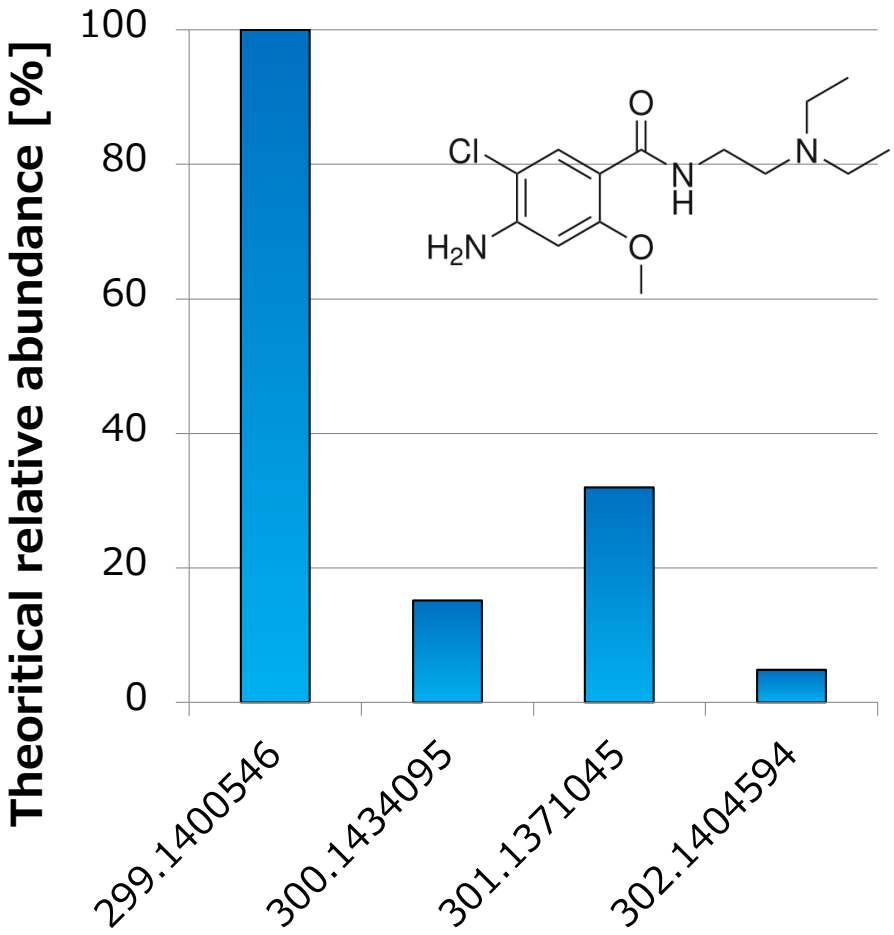
Lib: **b'** = (100, 1000, 100, 400, 300)

Rev. product: $\frac{a' \cdot b'}{|a'| \cdot |b'|} = 0.995$

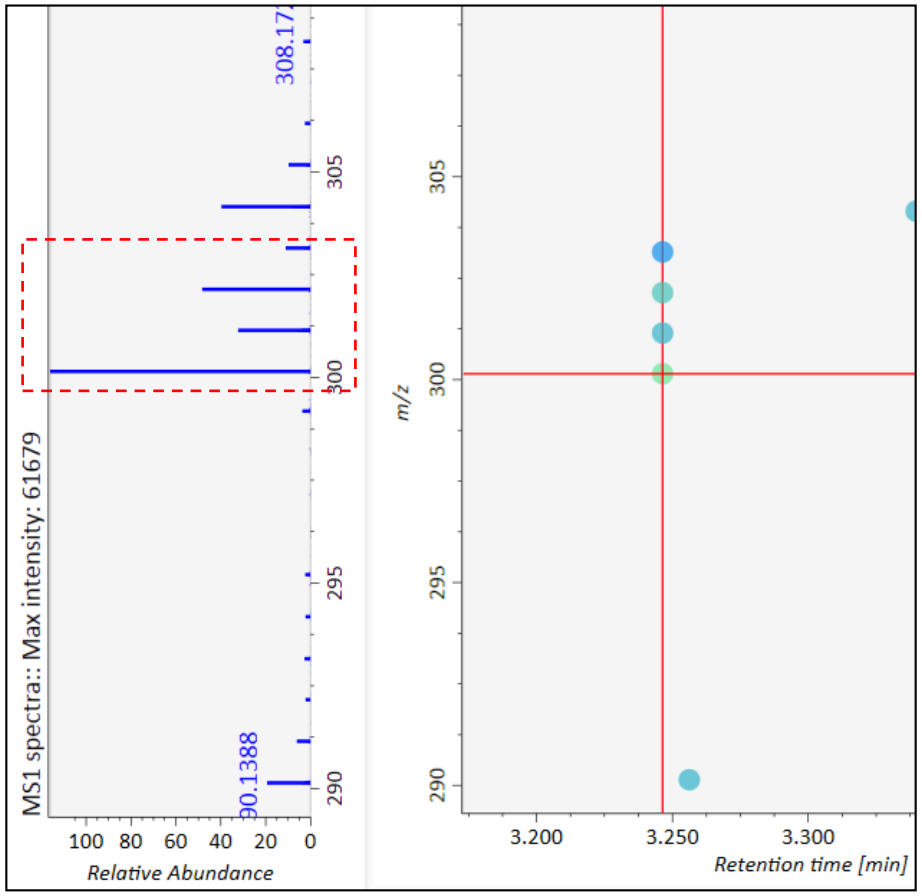
Presence ratio: $\frac{3}{5} = 0.6$

Isotope ratio similarity

Metoclopramide: $C_{14}H_{22}N_3O_2Cl$



Similarity: 0.915

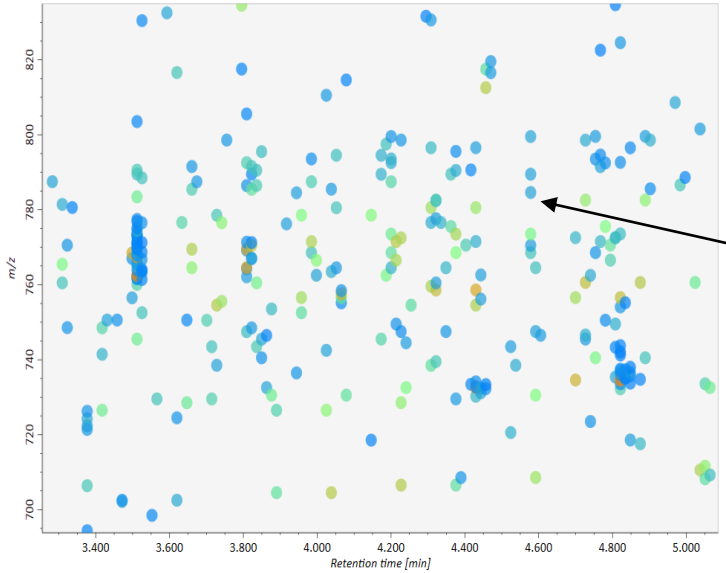


$$S_{ratio} = 1 - \sum |r_{act.i} - r_{lib.i}| \quad r_i = \frac{I_{M+i}}{I_M}, 1 \le i \le n$$

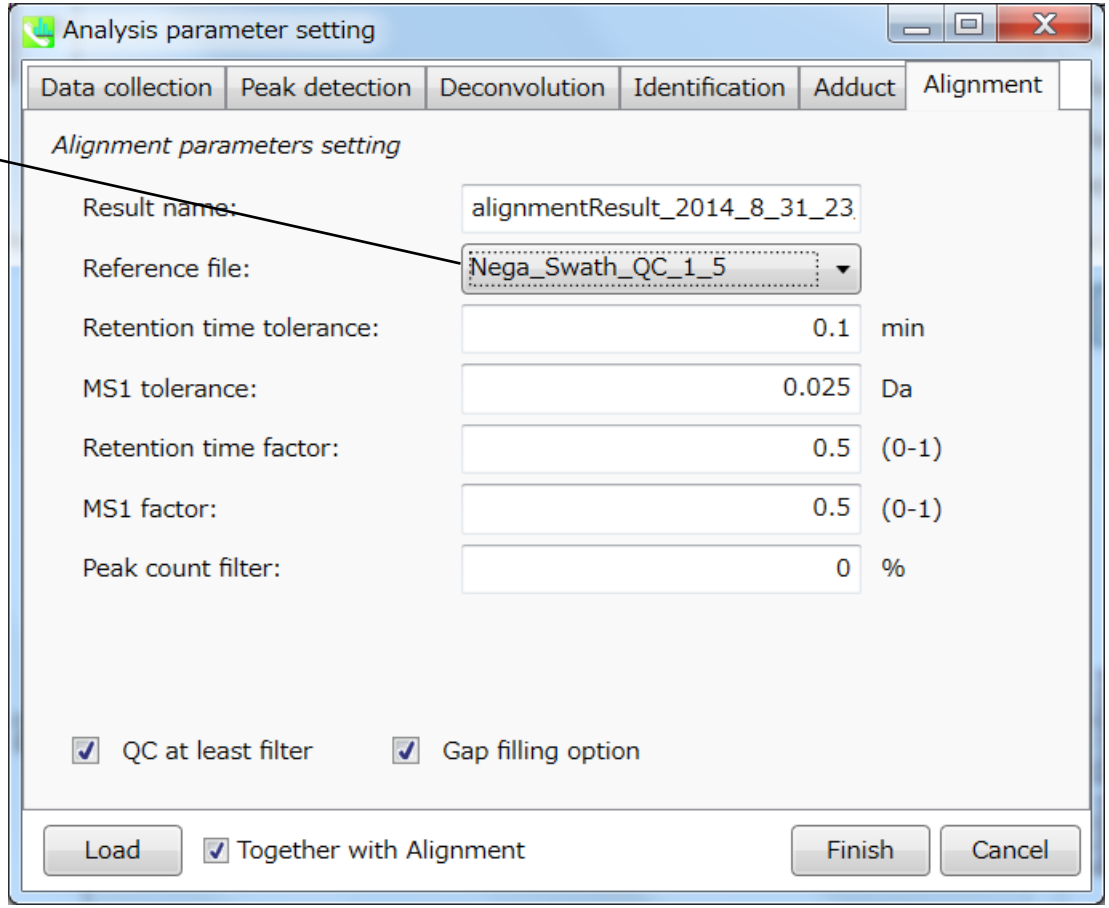
How it works for peak alignment

- ✓**Making ‘master peak list’**
- ✓**Joint aligner**
- ✓**Filtering of aligned peak list**
- ✓**Gap filing**

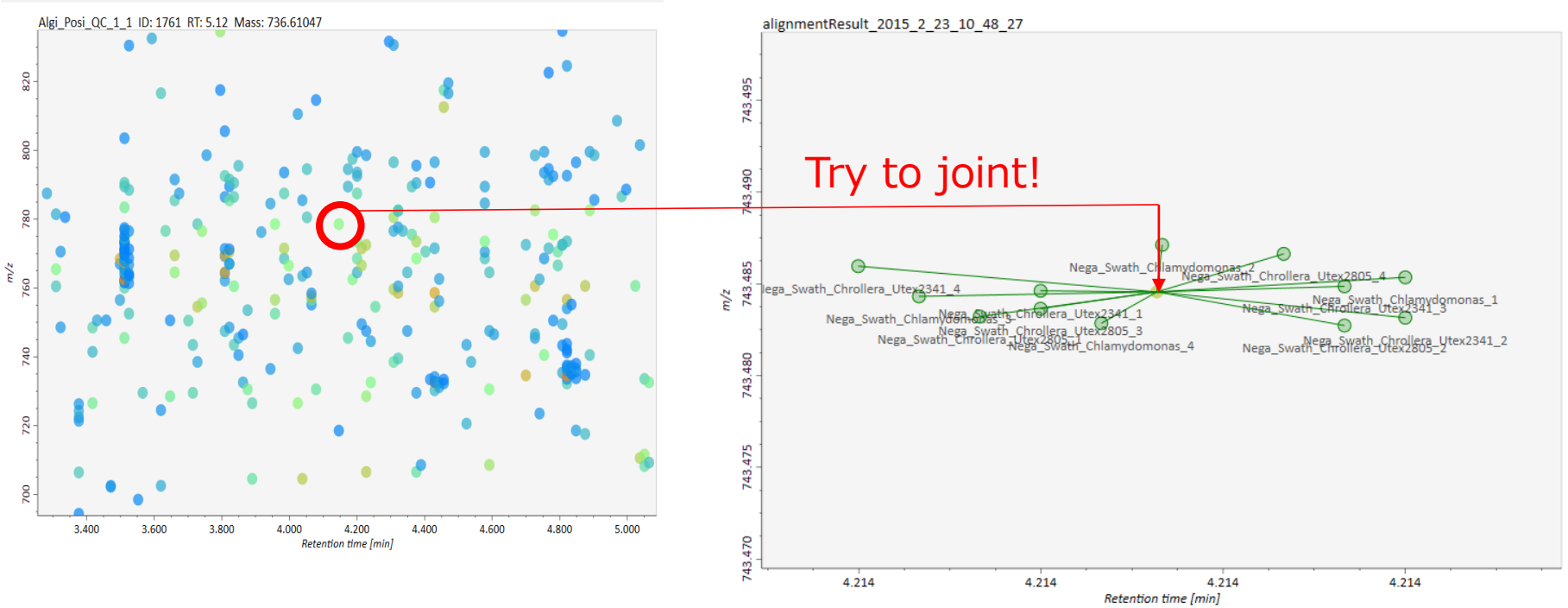
Making ‘master peak list’



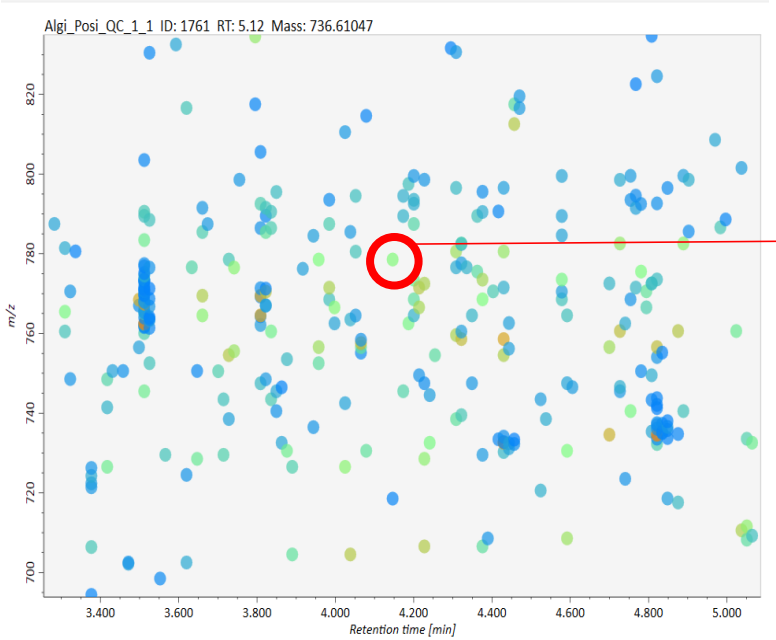
Peak no.	RT [min]	m/z
1	0.73	434.5876
2	1.26	842.1221
3	1.82	254.1078
4	2.11	332.0043
5	2.29	111.0078
k-1	7.09	982.9814
k	7.11	512.3321
k+1	7.12	687.8406
n	12.88	1230.412



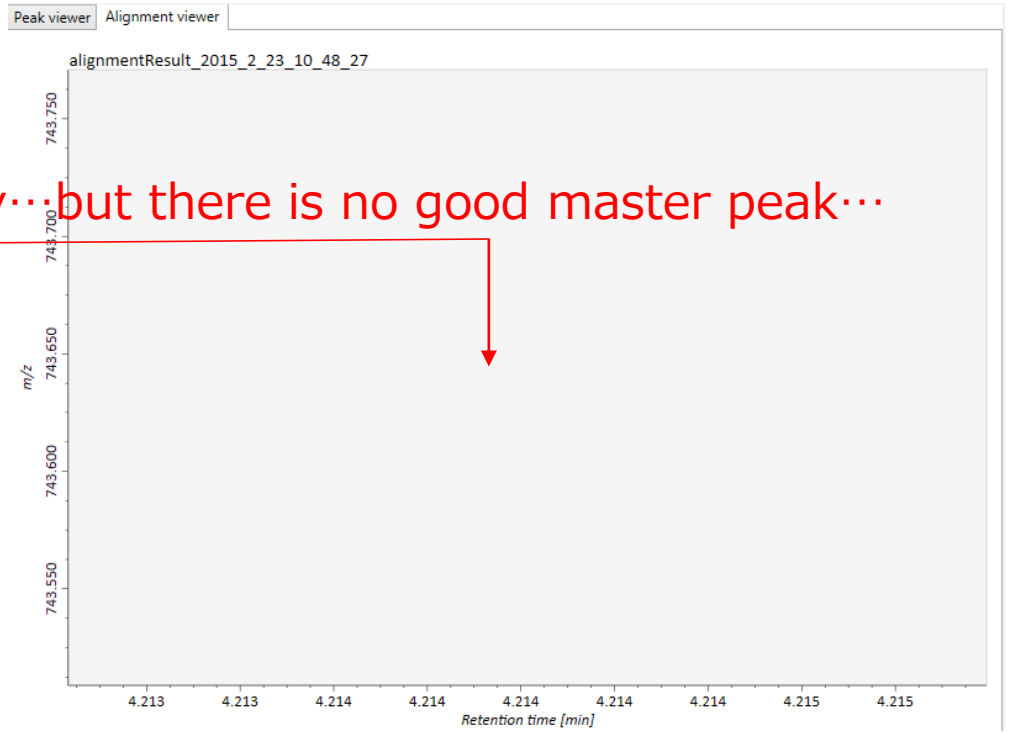
BTW, what is ‘Joint aligner’?



BTW 2, what if there is no master peak?

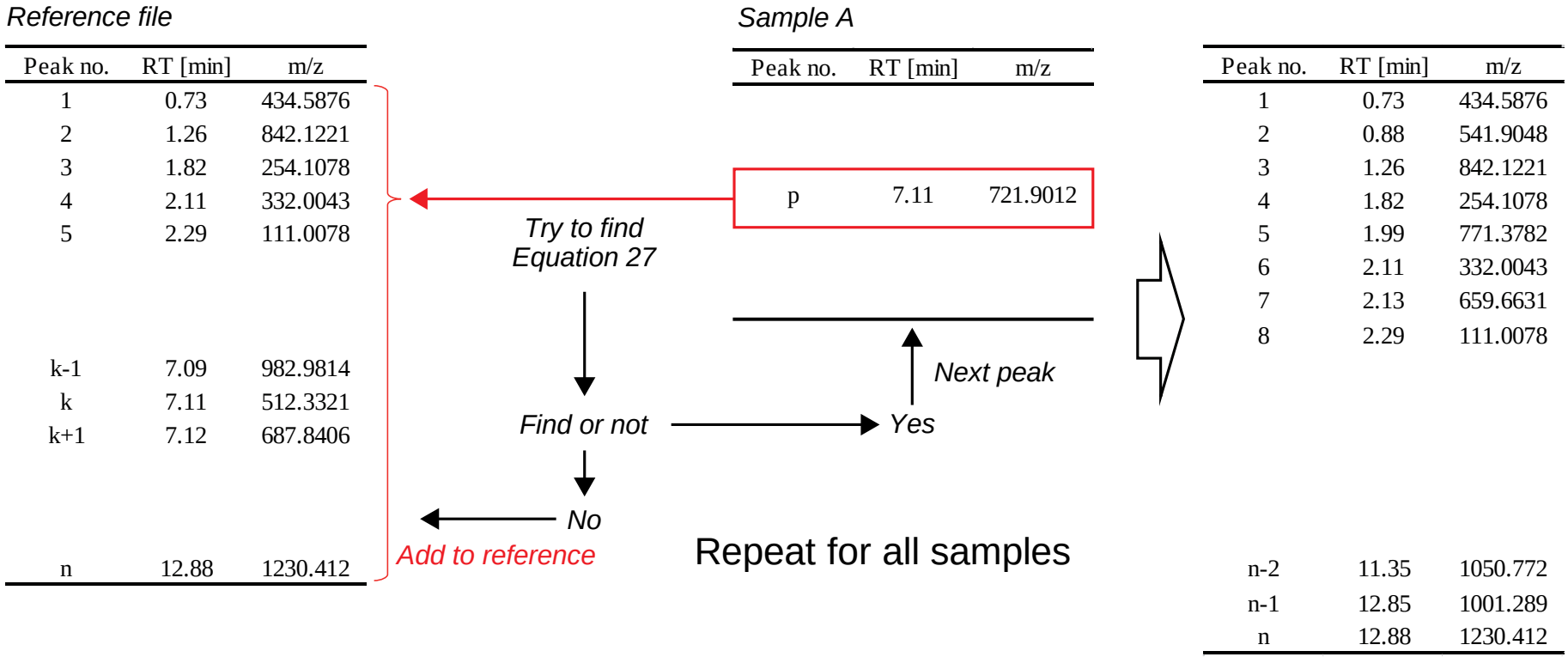


Try…but there is no good master peak…



Making ‘master peak list’

a. Making a reference peak table



Joint aligner

Reference peak table

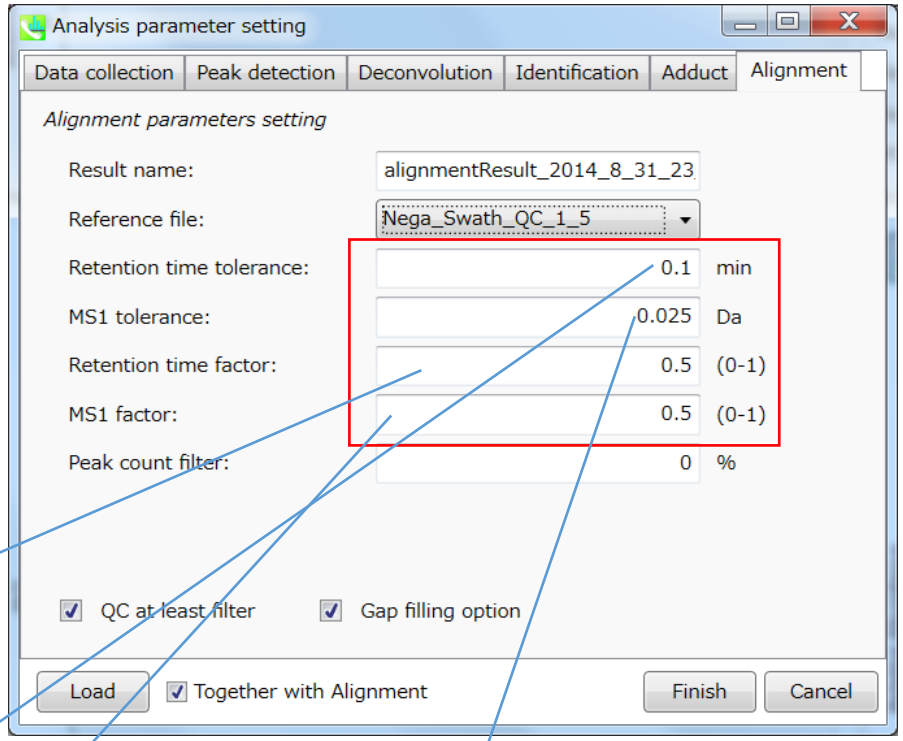
Peak no.	RT [min]	m/z
1	0.73	434.5876
2	0.88	541.9048
3	1.26	842.1221
4	1.82	254.1078
5	1.99	771.3782
6	2.11	332.0043
7	2.13	659.6631
8	2.29	111.0078
n-2	11.35	1050.772
n-1	12.85	1001.289
n	12.88	1230.412

Sample A

Peak no.	RT [min]	m/z
k	1.99	771.3761

Joint to matched peak
Equation 28

Repeat for all samples



$$\text{Score} = a \times \exp\left\{-0.5 \times \left(\frac{RT_{sam.} - RT_{ref.}}{\delta_{RT}}\right)^2\right\} + b \times \exp\left\{-0.5 \times \left(\frac{Mass_{sam.} - Mass_{ref.}}{\delta_{Mass}}\right)^2\right\} \quad \text{(eq. 28)}$$

Joint aligner

b. Fitting each sample peak table to reference peak table

Reference peak table

Peak no.	RT [min]	m/z
1	0.73	434.5876
2	0.88	541.9048
3	1.26	842.1221
4	1.82	254.1078
5	1.99	771.3782
6	2.11	332.0043
7	2.13	659.6631
8	2.29	111.0078
n-2	11.35	1050.772
n-1	12.85	1001.289
n	12.88	1230.412

Sample A

Peak no.	RT [min]	m/z
k	1.99	771.3761

Joint to matched peak
Equation 28

Repeat for all samples

Aligned peak table

Alignment ID	RT Ave	m/z Ave.	Sample A	Sample B	Sample C
1	0.72	434.5878	2500	1800	4000
2	0.88	541.9050	N.D.	1500	2000
3	1.25	842.1220	53000	62000	40000
4	1.81	254.1079	100	50	730
5	1.99	771.3765	N.D.	N.D.	N.D.
6	2.12	332.0049	14500	7800	25000
7	2.13	659.6631	90000	150000	120000
8	2.28	111.0082	8500	N.D.	N.D.
n-2	11.36	1050.775	5000	4500	N.D.
n-1	12.85	1001.282	58000	20000	25000
n	12.87	1230.412	10000	12000	9000

Filtering of aligned peak list

c. Filtering aligned peaks

Aligned peak table

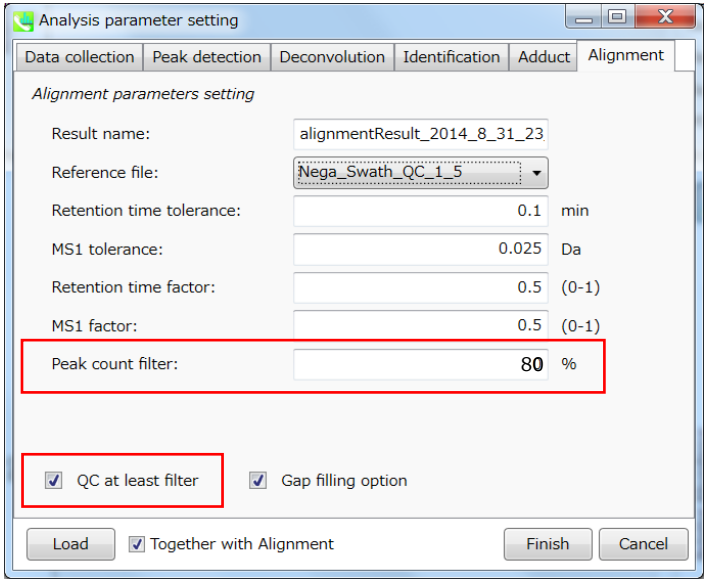
Alignment ID	RT Ave.	m/z Ave.	Sample A	Sample B	Sample C	QC 1	QC 2	QC 3
1	0.72	434.5878	2500	1800	4000	1000	N.D.	2000
2	0.88	541.9050	N.D.	1500	2000	1400	2000	1500
3	1.25	842.1220	53000	62000	40000	45000	30000	35000
4	1.81	254.1079	100	50	730	100	50	730
5	1.99	771.3765	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
6	2.12	332.0049	14500	7800	25000	14500	7800	25000
7	2.13	659.6631	90000	150000	120000	75000	70000	72000
8	2.28	111.0082	8500	N.D.	N.D.	8500	8800	9000
n	12.87	1230.412	10000	12000	9000	15000	12000	13000

Condition
1. Peak count filter: 80%
2. QC 'at least' filter: ON

← *Excluded (Step 3)*

← *Excluded (Step 1)*

← *Excluded (Step 2)*



Gap filing: interpolating missing values

d. Interpolating missing values

The maximum of raw data point within the blue range (equation 29) is interpolated.

Alignment ID	RT Ave.	m/z Ave.	Sample A	Sample B	Sample C	QC 1	QC 2	QC 3
1'	0.88	541.9050	N.D.	1500	2000	1400	2000	1500
2'	1.25	842.1220	53000	62000	40000	45000	30000	35000
3'	1.81	254.1079	100	50	730	100	50	730
4'	2.11	332.0049	14500	7800	25000	14500	7800	25000
5'	2.13	659.6631	90000	150000	120000	75000	70000	72000
n'	12.87	1230.412	10000	12000	9000	15000	12000	13000