

MS-DIAL FAQ

Hiroshi Tsugawa



How it works for peak detections

✓ **Differential function**

✓ **Noise estimation**

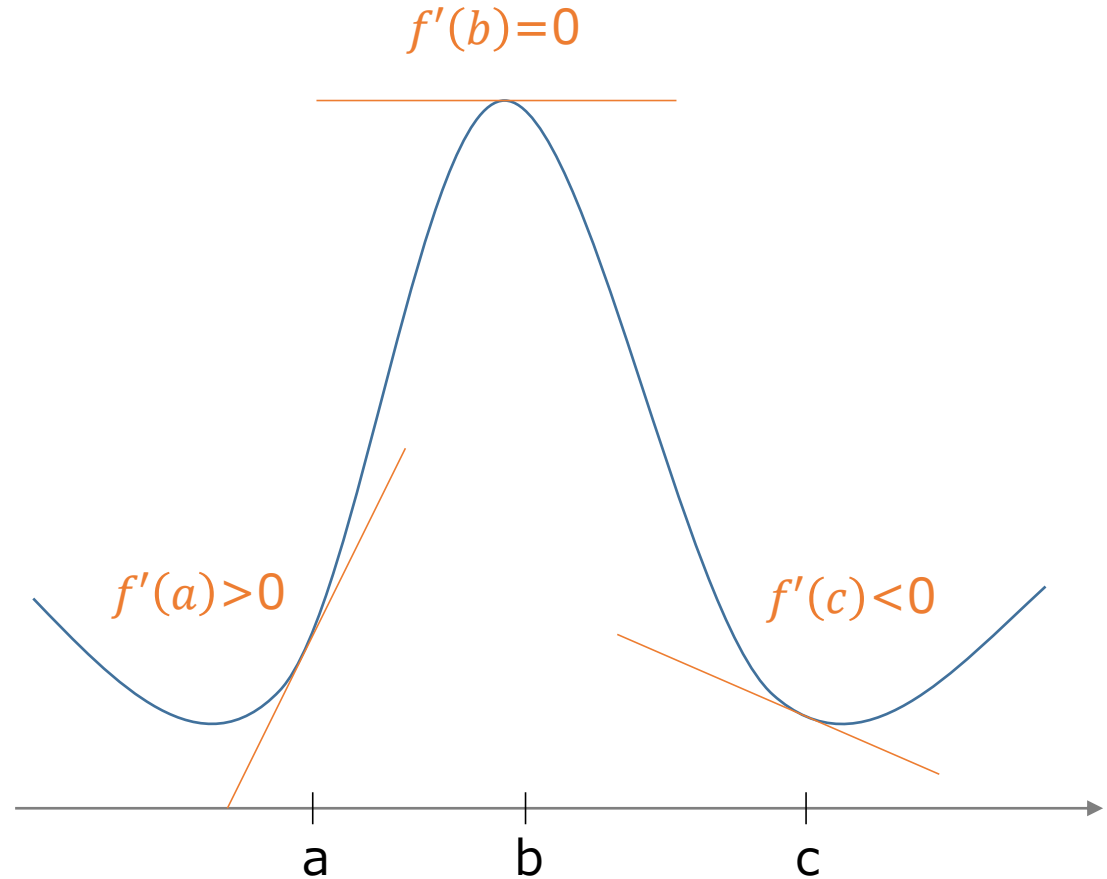
Differential function

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



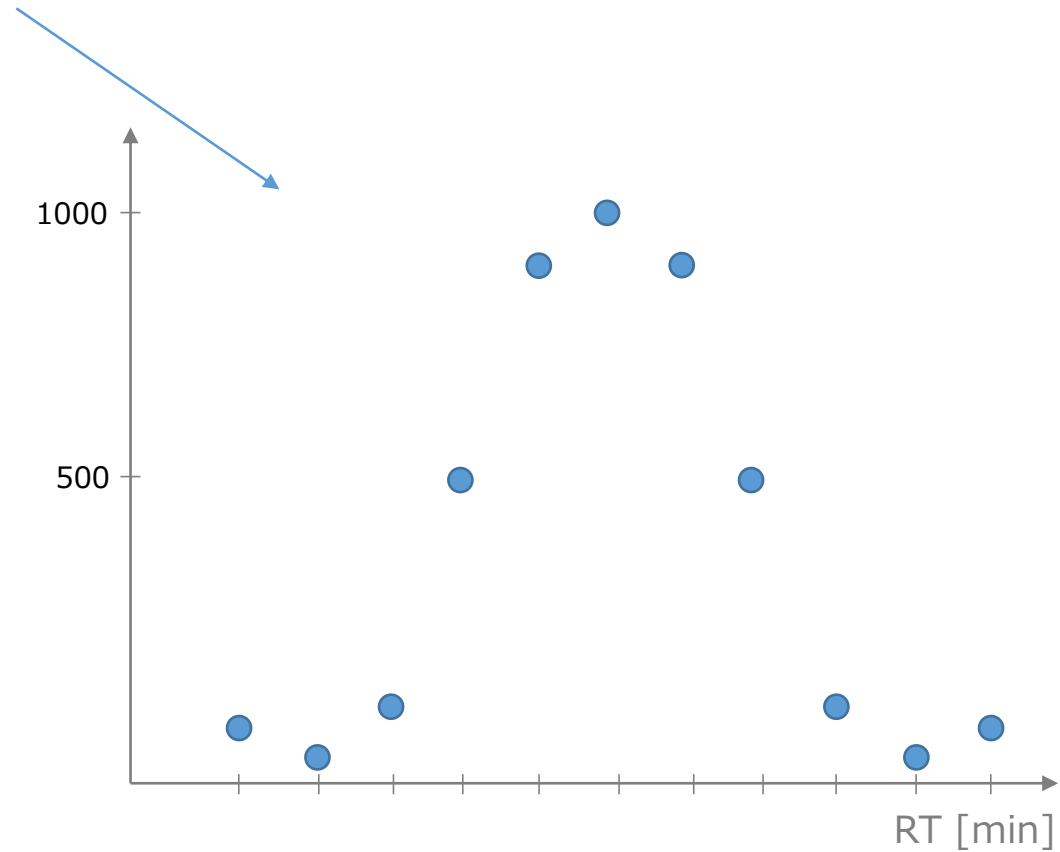
$$f'(x) = 3ax^2 + 2bx + c$$

Slope



Differential function for chromatogram

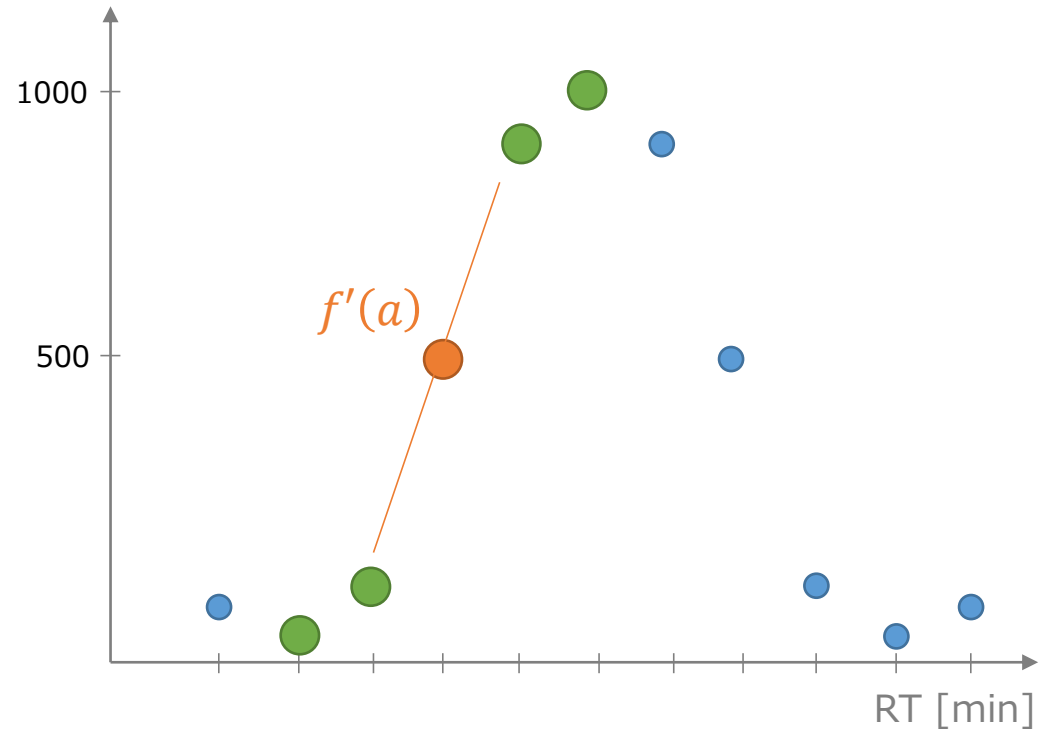
$$f(x) = \{50, 10, 100, 500, 800, 1000, 800, 500, 100, 10, 50\}$$



Differential function for chromatogram

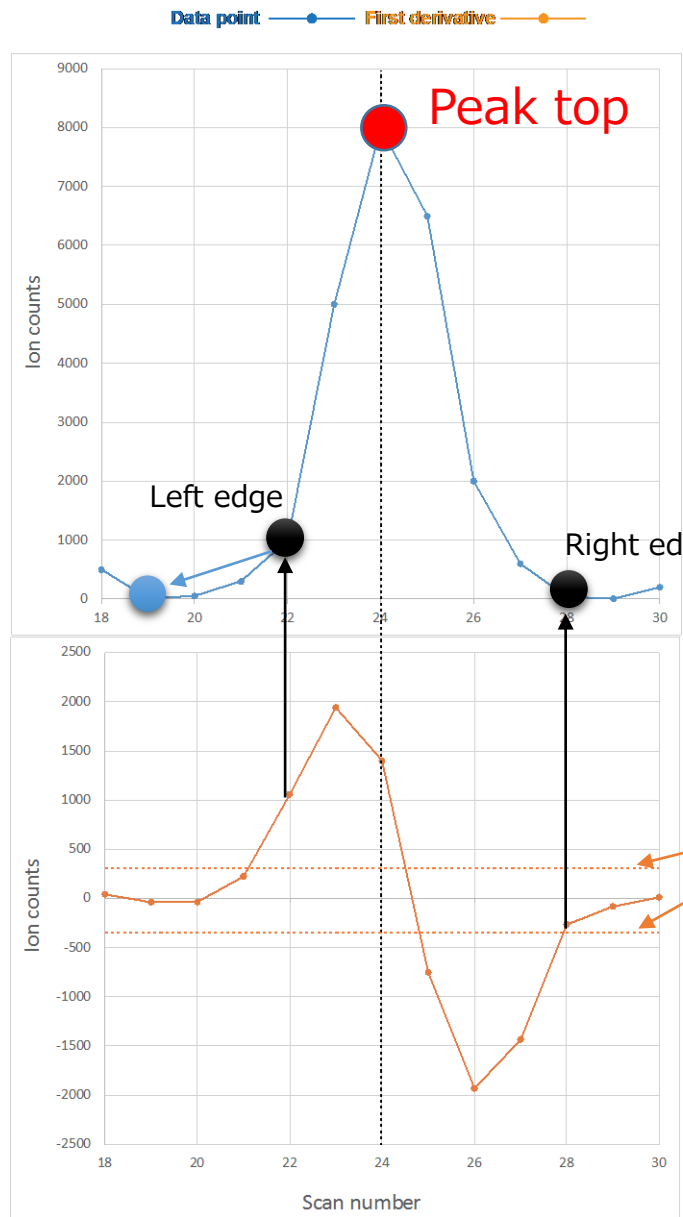
$$f(x) = \{50, 10, 100, 500, 800, 1000, 800, 500, 100, 10, 50\}$$

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



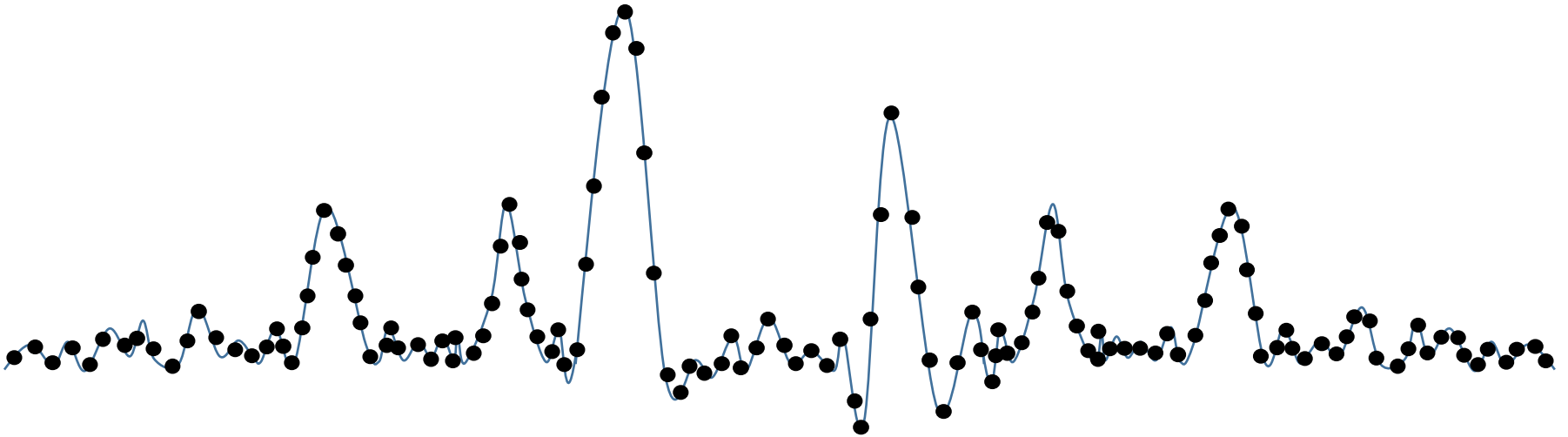
Differential function for chromatogram

Back tracing



Noise threshold

Noise evaluation



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

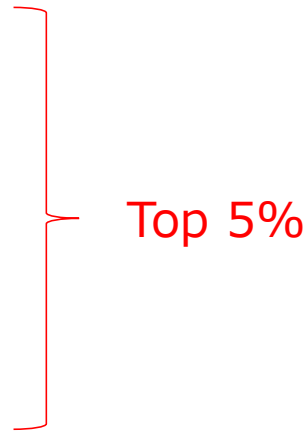


Sort the first derivative array by the absolute values

Noise evaluation

Noise value = Median

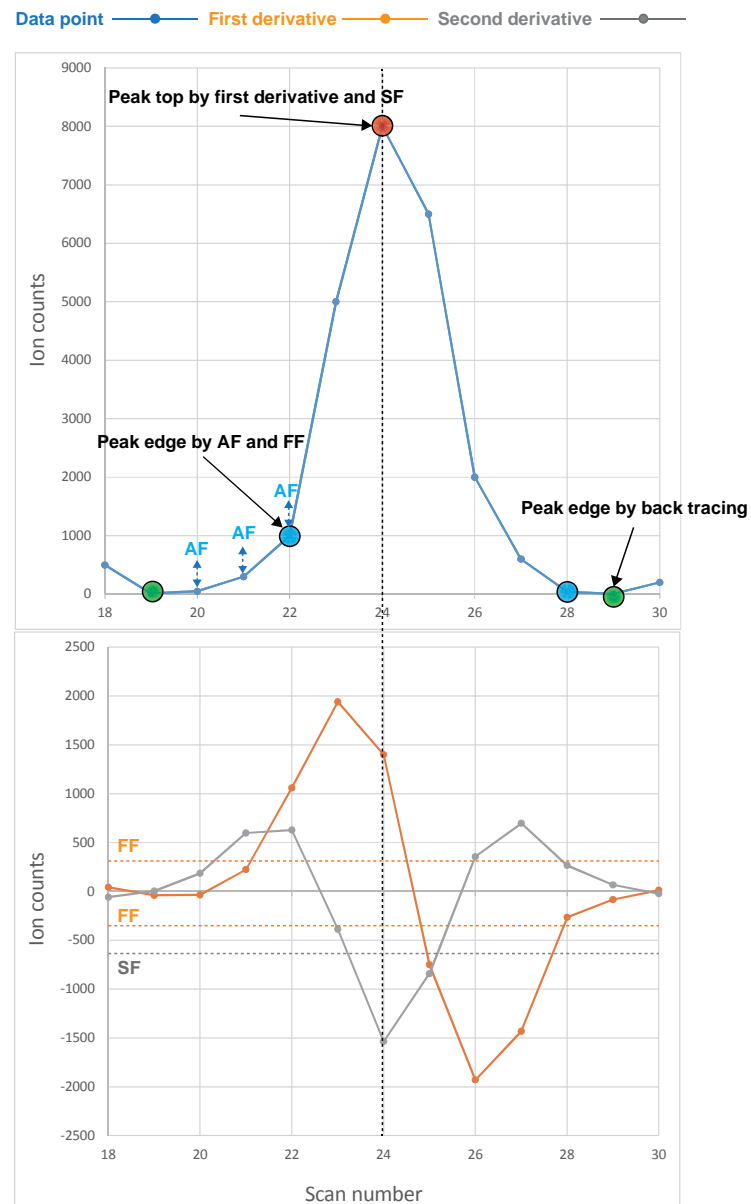
<u>f'(x)</u>
0.361563
2.368109
6.641594
9.488878
10.547
11.025
16.82708
17.22223
21.78178
22.57424
23.91914
27.35502
30.6929
31.75519
32.60672
36.23518
40.57042
45.80185
47.27034
55.0006
58.68845
58.71436
61.12281
61.28744
70.8877
73.31488
77.39816
77.75772
77.99117
80.14925
85.77214
86.85782
•
•
•



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

Analysis parameter setting

Data collection | Peak detection | Deconvolution | Identification | Adduct | Alignment

Peak detection parameters

Smoothing method: Linear weighted moving aver. ▾

Smoothing level: 2 scan

Minimum peak width: 5 scan

Minimum peak height: 500 amplitude

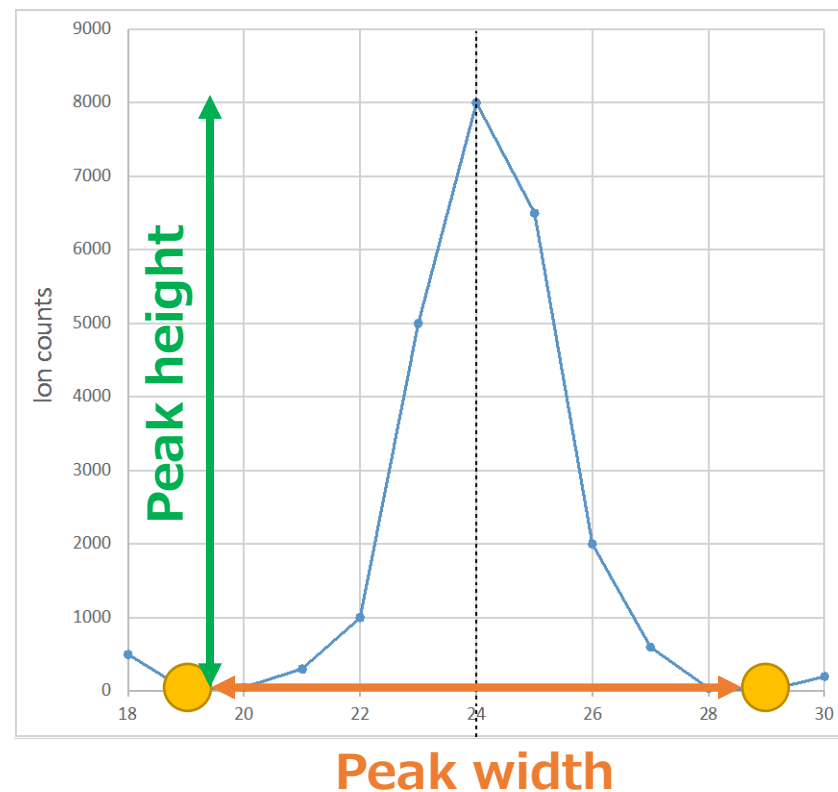
Peak spotting parameters

Mass slice width: 0.1 Da

Exclusion mass list:

Accurate mass [Da]	Mass tolerance [Da]
100.9342	0.005
112.9862	0.005
115.9207	0.005
130.9668	0.005
132.9243	0.005
139.0088	0.005
158.9769	0.005
182.9895	0.005
186.9327	0.005
212.0751	0.005
213.0784	0.005
248.9622	0.005
248.9739	0.005
316.9489	0.005
384.9353	0.005

Load Together with Alignment Finish Cancel



In MS-DIAL program

Analysis parameter setting

Data collection Peak detection Deconvolution Identification Adduct Alignment

Peak detection parameters

Smoothing method: Linear weighted moving aver. ▼

Smoothing level: 2 scan

Minimum peak width: 5 scan

Minimum peak height: 500 amplitude

Peak spotting parameters

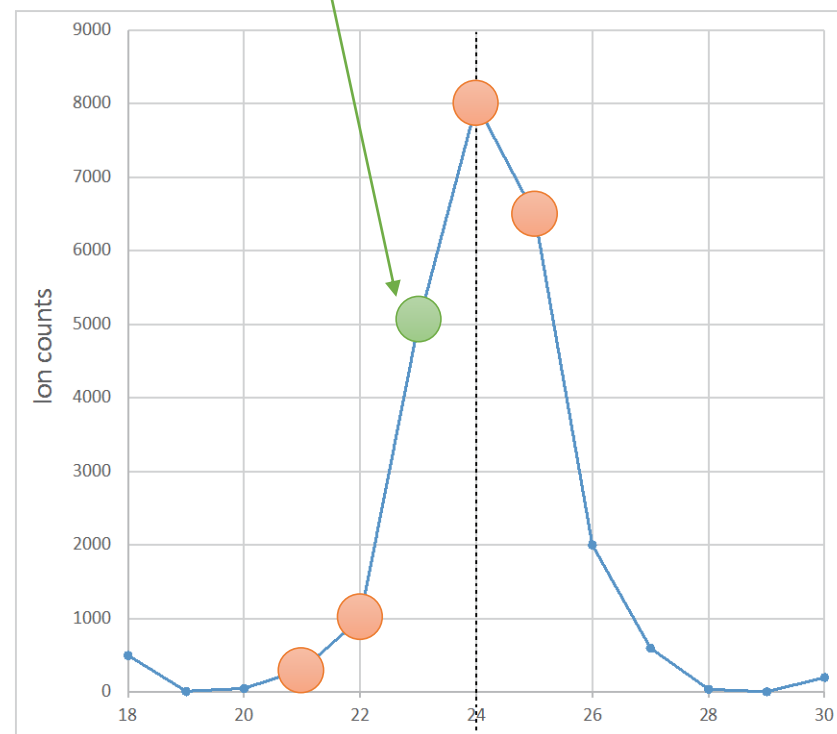
Mass slice width: 0.1 Da

Exclusion mass list:

Accurate mass [Da]	Mass tolerance [Da]
100.9342	0.005
112.9862	0.005
115.9207	0.005
130.9668	0.005
132.9243	0.005
139.0088	0.005
158.9769	0.005
182.9895	0.005
186.9327	0.005
212.0751	0.005
213.0784	0.005
248.9622	0.005
248.9739	0.005
316.9489	0.005
384.9353	0.005

Load Together with Alignment Finish Cancel

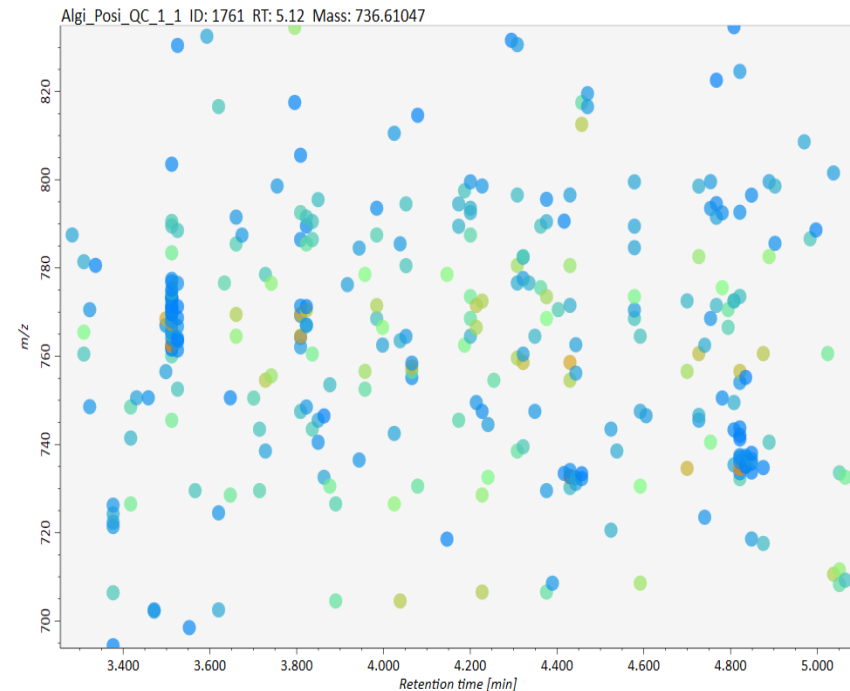
$$\text{smoothed value} = \frac{1f(x_{-2}) + 2f(x_{-1}) + 3f(x) + 2f(x_{+1}) + 1f(x_{+2})}{9}$$



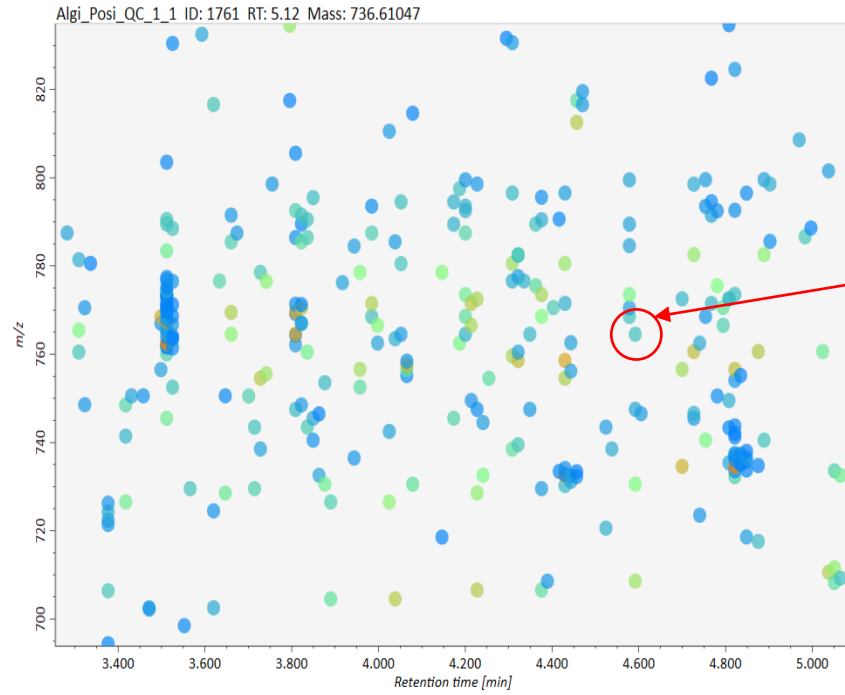
How it works in RT & m/z axis

✓ Slicing method

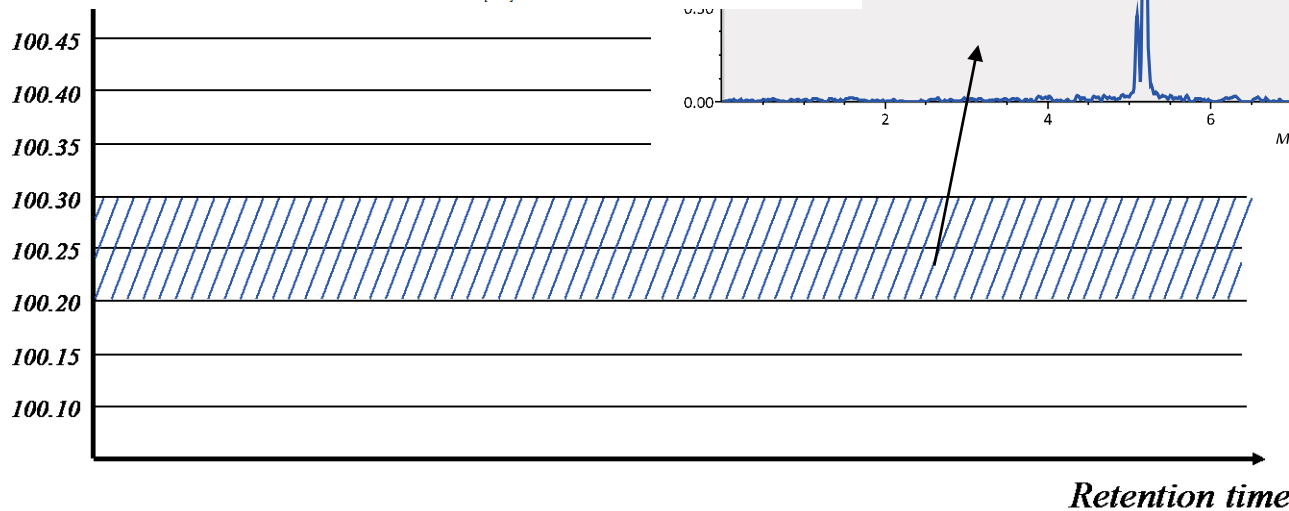
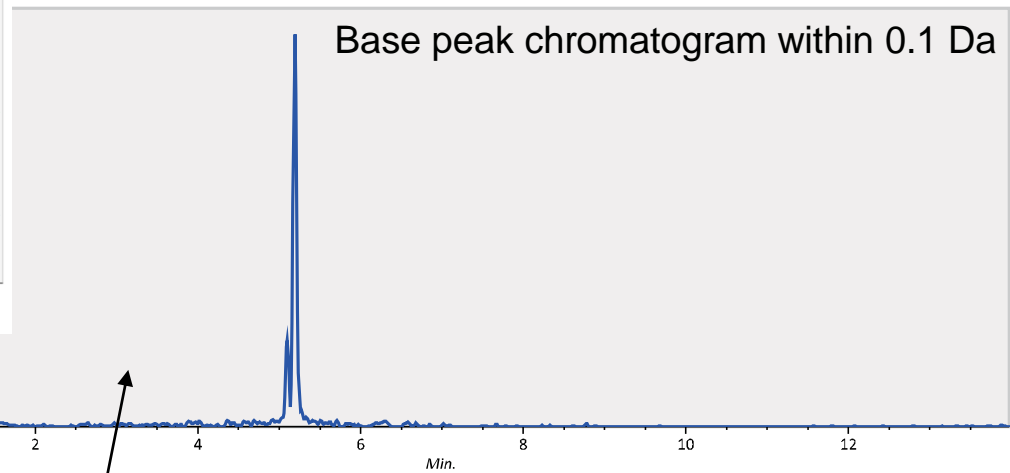
✓ Peak detection is applied to
base peak chromatogram



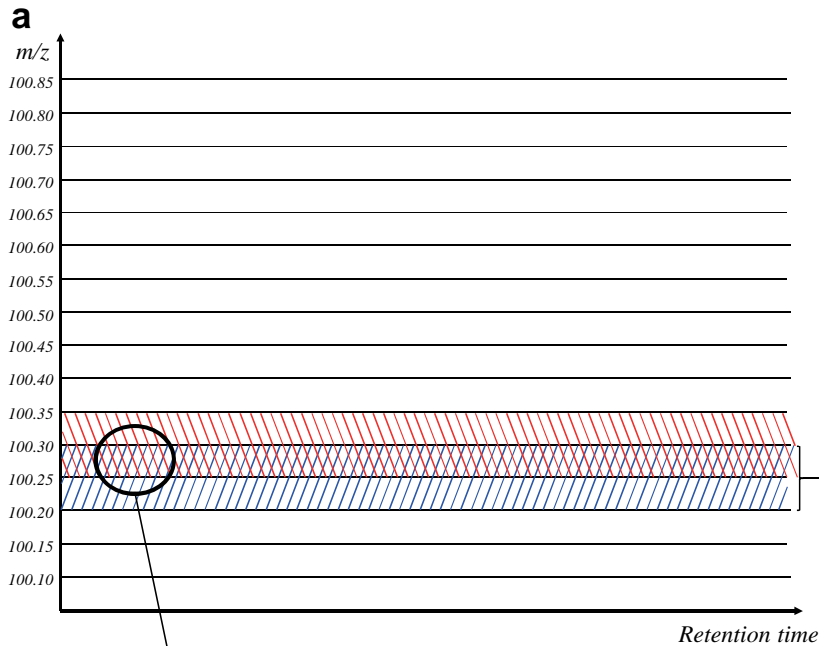
Slicing method & BPC extraction



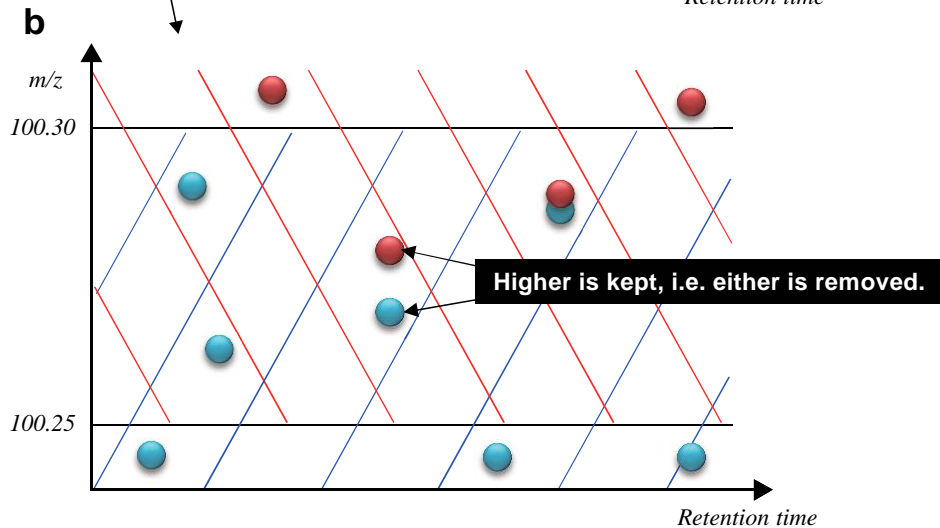
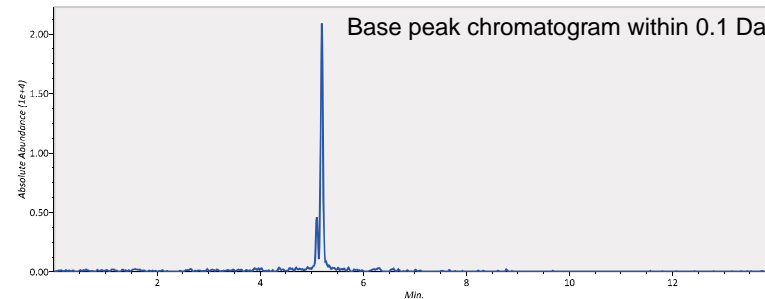
Scan number	Retention time [min]	Base peak m/z	Base peak intensity
1	0.1	100.2054	1
2	0.12	100.2053	10
3	0.14	100.2053	5
4	0.16	100.2052	50
5	0.18	100.2051	200
6	0.2	100.2054	1500
7	0.22	100.2054	3000
8	0.24	100.2054	1700
9	0.26	100.2053	180
10	0.28	100.205	60



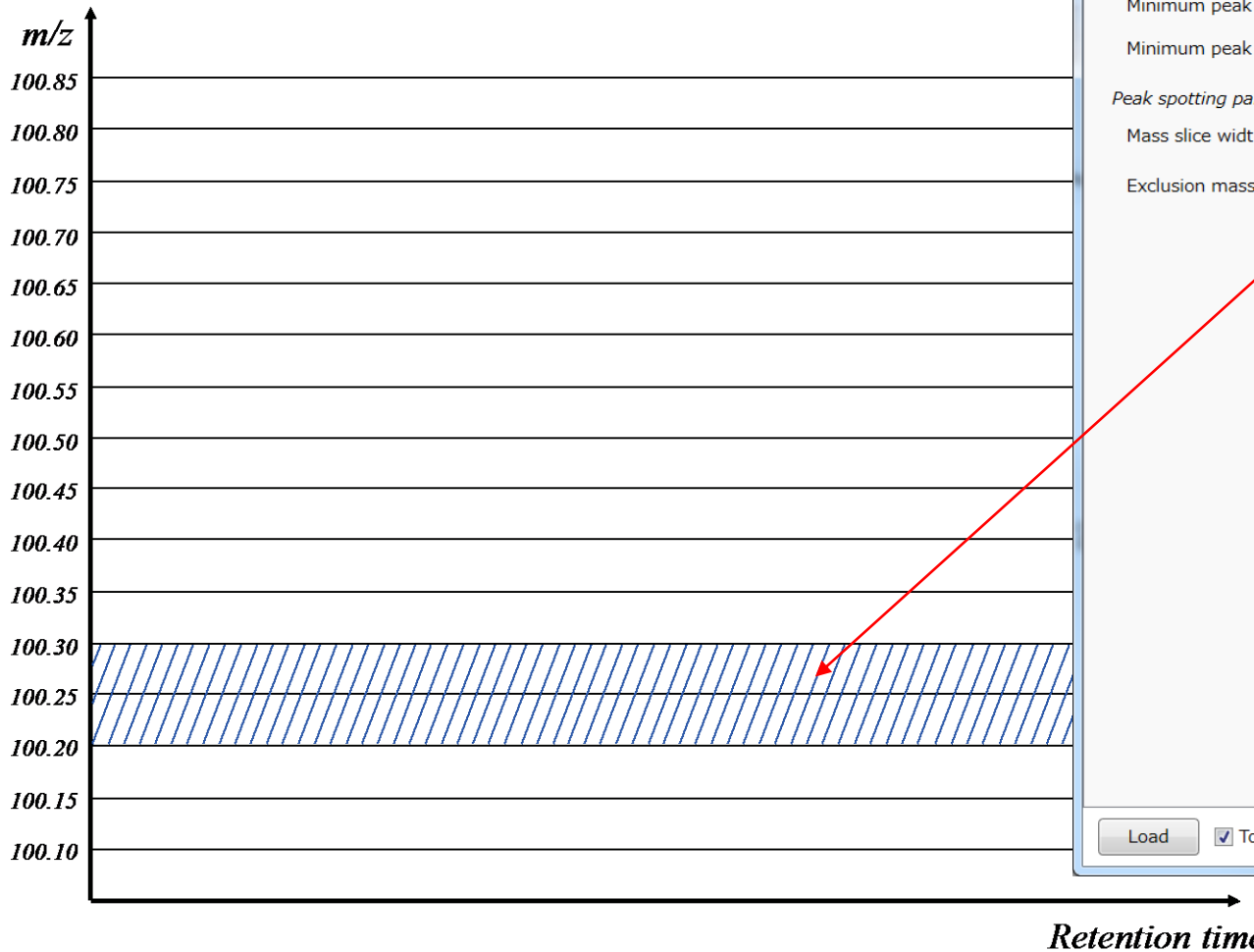
Slicing method & BPC extraction



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



Analysis parameter setting

Data collection Peak detection Deconvolution Identification Adduct Alignment

Peak detection parameters

Smoothing method: Linear weighted moving aver

Smoothing level: 2 scan

Minimum peak width: 5 scan

Minimum peak height: 500 amplitude

Peak spotting parameters

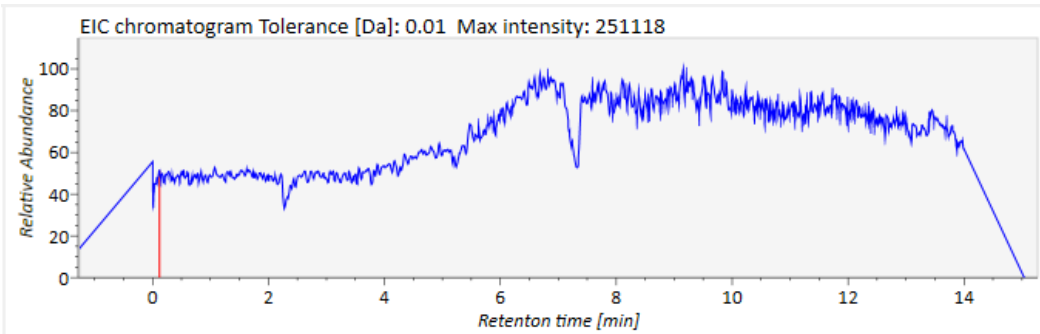
Mass slice width: 0.1 Da

Exclusion mass list:

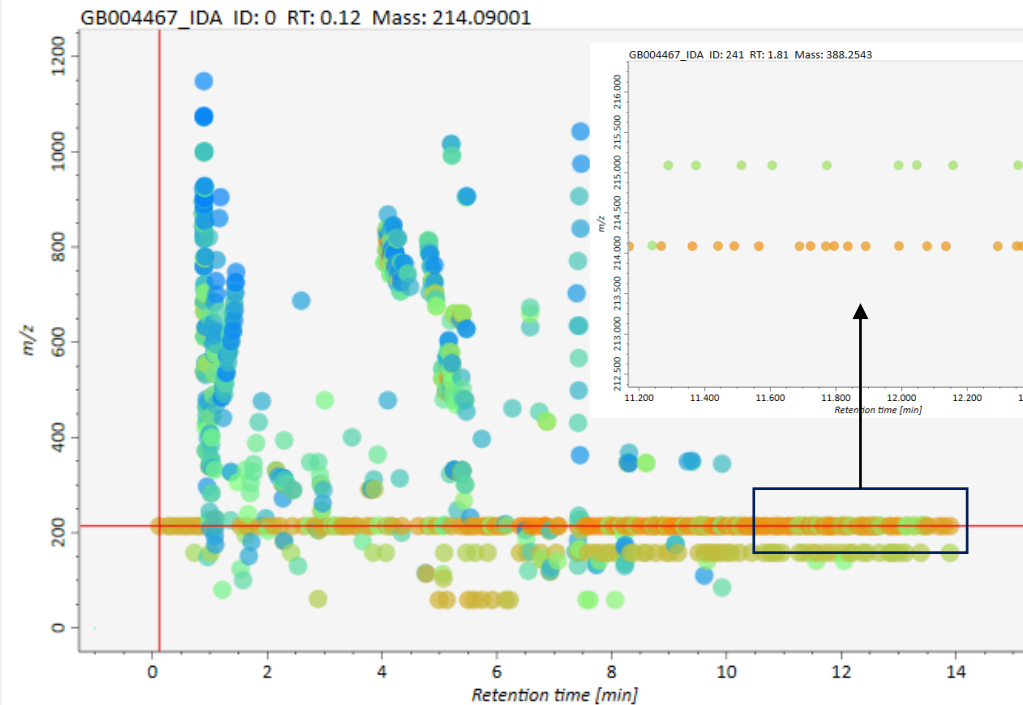
Accurate mass [Da]	Mass tolerance [Da]
100.9342	0.005
112.9862	0.005
115.9207	0.005
130.9668	0.005
132.9243	0.005
139.0088	0.005
158.9769	0.005
182.9895	0.005
186.9327	0.005
212.0751	0.005
213.0784	0.005
248.9622	0.005
248.9739	0.005
316.9489	0.005
384.9353	0.005

Load Together with Alignment Finish Cancel

What is 'exclusion mass list'?



Peak viewer Alignment viewer



Analysis parameter setting

Data collection Peak detection Deconvolution Identification Adduct Alignment

Peak detection parameters

Smoothing method: Linear weighted moving aver

Smoothing level: 2 scan

Minimum peak width: 5 scan

Minimum peak height: 500 amplitude

Peak spotting parameters

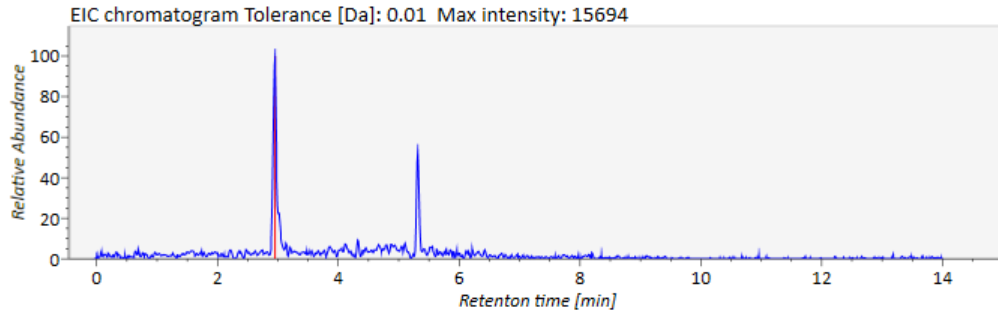
Mass slice width: 0.1 Da

Exclusion mass list:

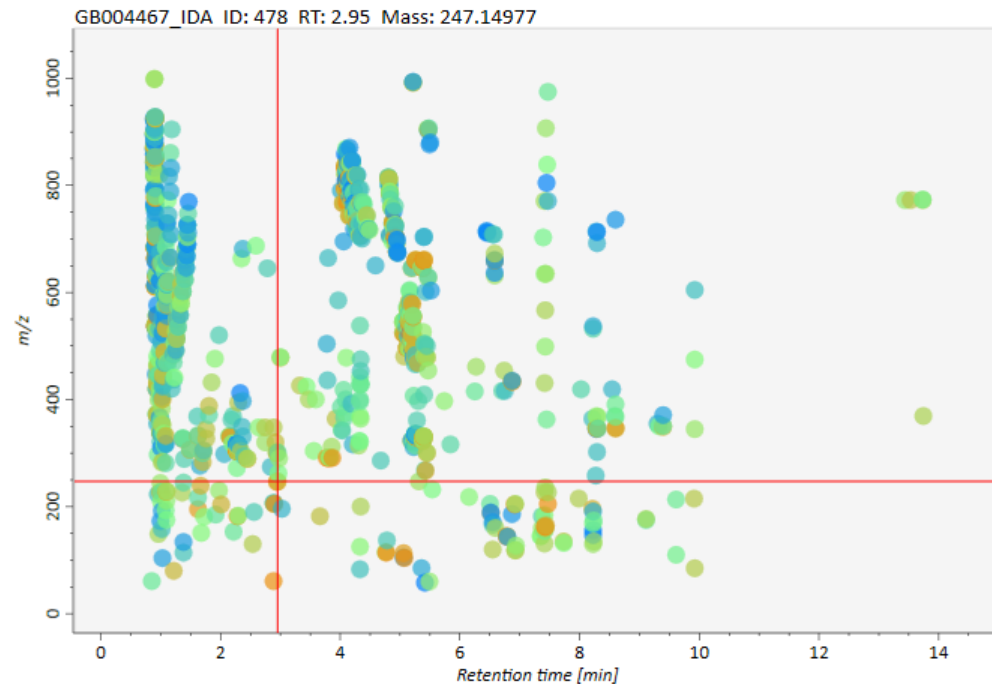
Accurate mass [Da]	Mass tolerance [Da]
100.9342	0.005
112.9862	0.005
115.9207	0.005
130.9668	0.005
132.9243	0.005
139.0088	0.005
158.9769	0.005
182.9895	0.005
186.9327	0.005
212.0751	0.005
213.0784	0.005
248.9622	0.005
248.9739	0.005
316.9489	0.005
384.9353	0.005

Load Together with Alignment Finish Cancel

What is 'exclusion mass list'?



Peak viewer Alignment viewer



Analysis parameter setting

Data collection Peak detection Deconvolution Identification Adduct Alignment

Peak detection parameters

Smoothing method: Linear weighted moving aver

Smoothing level: 2 scan

Minimum peak width: 5 scan

Minimum peak height: 500 amplitude

Peak spotting parameters

Mass slice width: 0.1 Da

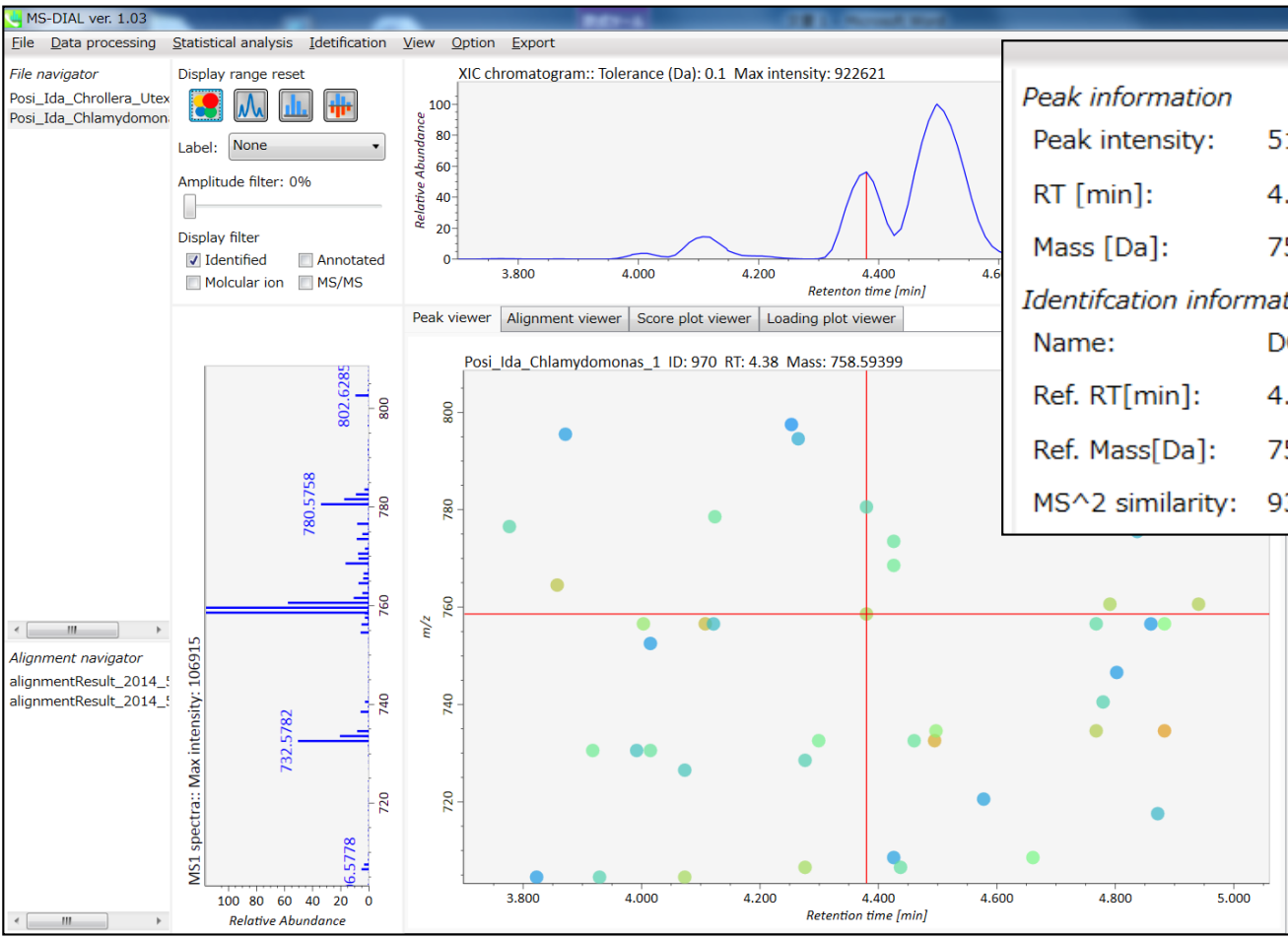
Exclusion mass list:

Accurate mass [Da]	Mass tolerance [Da]
100.9342	0.005
112.9862	0.005
115.9207	0.005
130.9668	0.005
132.9243	0.005
139.0088	0.005
158.9769	0.005
182.9895	0.005
186.9327	0.005
212.0751	0.005
213.0784	0.005
248.9622	0.005
248.9739	0.005
316.9489	0.005
384.9353	0.005

Load Together with Alignment Finish Cancel

Retention time and MS1 similarity

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



Peak information

- Peak intensity: 519303
- RT [min]: 4.38
- Mass [Da]: 758.594

Identification information

- Name: DGTS(18:2/18:3); [M+H]
- Ref. RT[min]: 4.631
- Ref. Mass[Da]: 758.5934 (Isotope: 901)
- MS² similarity: 936 (Reverse: 999)

Annotations:

- RT difference: 0.251 min (indicated by a red arrow between 4.38 and 4.631)
- Mass difference: 0.0006 Da (indicated by a blue arrow between 758.594 and 758.5934)

Retention time and MS1 similarity

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

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

Analysis parameter setting

Data collection | Peak detection | Deconvolution | Identification | Adduct | Alignment

MSP file and MS/MS identification setting

MSP file: F:\140421_Ida_Nega_AlgaesDatabase\LipidBlast_Nega_Ali [Select]

Retention time tolerance: 0.5 min

Accurate mass tolerance (MS1): 0.025 Da

Accurate mass tolerance (MS2): 0.25 Da

Identification score cut off: 70 %

Text file and post identification (retention time and accurate mass based) setting

Text file: F:\20140809_MSIAL_DemoFiles_Swath (abf)\Lipid_Nega [Select]

Retention time tolerance: 0.25 min

Accurate mass tolerance: 0.01 Da

Identification score cut off: 85 %

Advanced library search option

Load Together with Alignment Finish Cancel

Peak information

Peak intensity: 519303

RT [min]: 4.38

Mass [Da]: 758.594

Identification information

Name: DGTS(18:2/18:3); [M+H]

Ref. RT [min]: 4.631

Ref. Mass [Da]: 758.5934 (Isotope: 901)

MS² similarity: 936 (Reverse: 999)

RT difference: 0.251 min

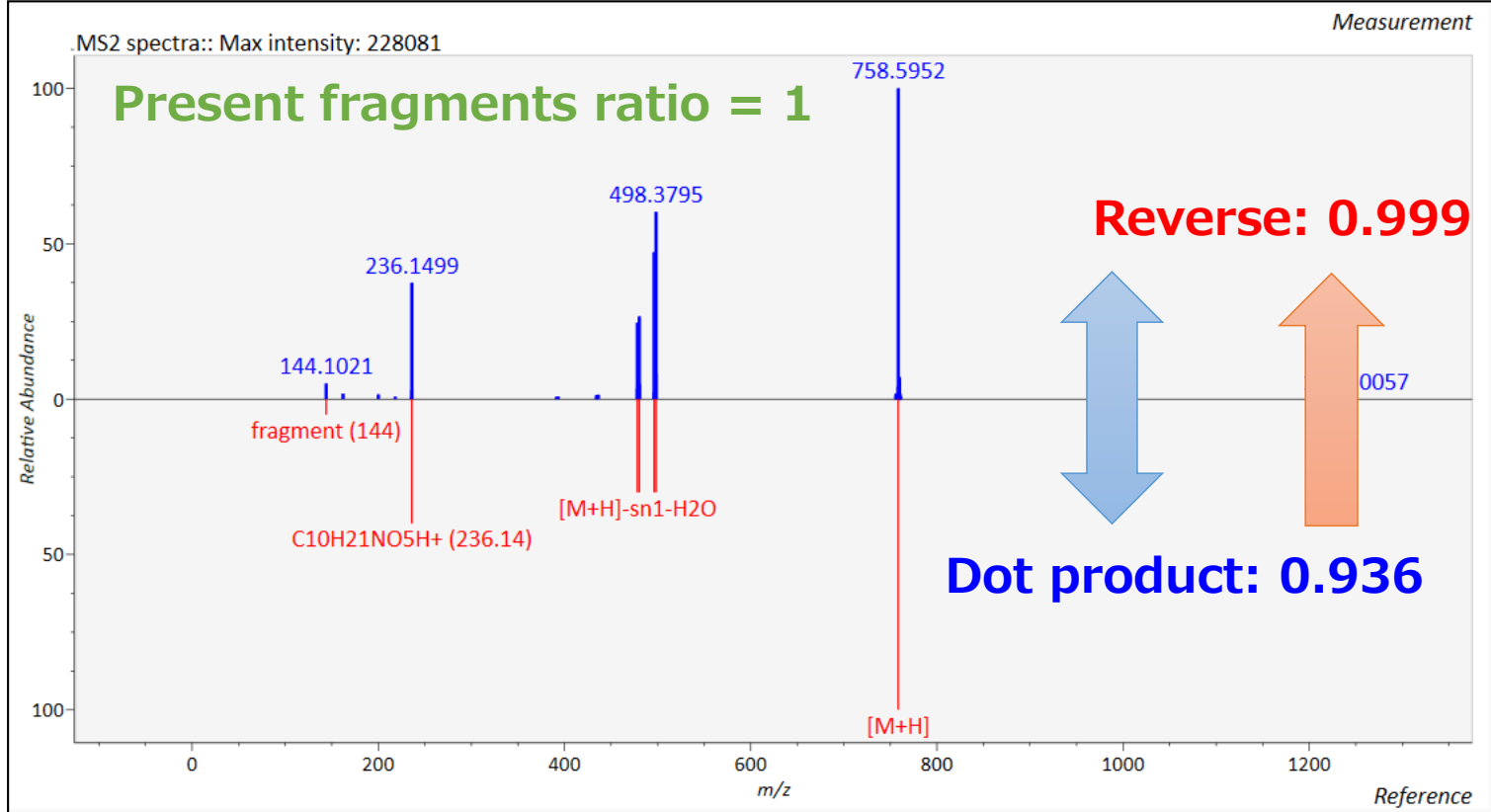
Mass difference: 0.0006 Da

MS/MS similarity

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

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

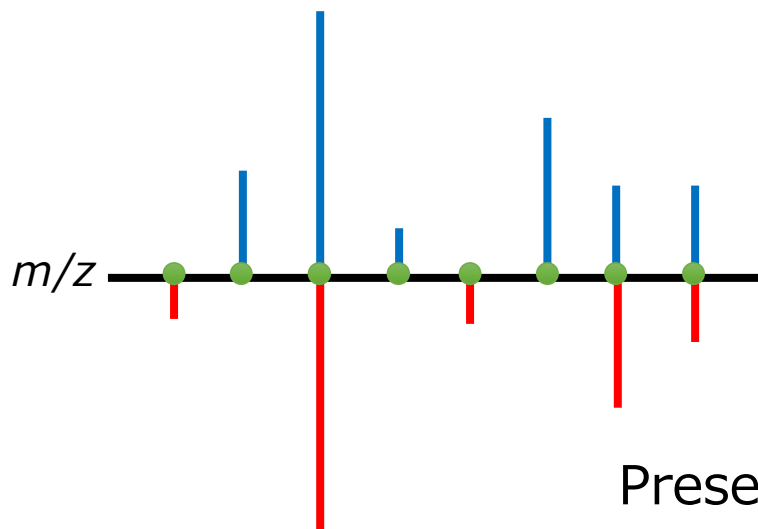
$$Present\ fragments\ ratio = \frac{\text{The number of mated fragments}}{\text{The number of reference fragments}}$$



BTW, what is the difference?

Dot product

Act.: $\mathbf{a} = (0, 400, 1000, 200, 0, 700, 300, 300)$

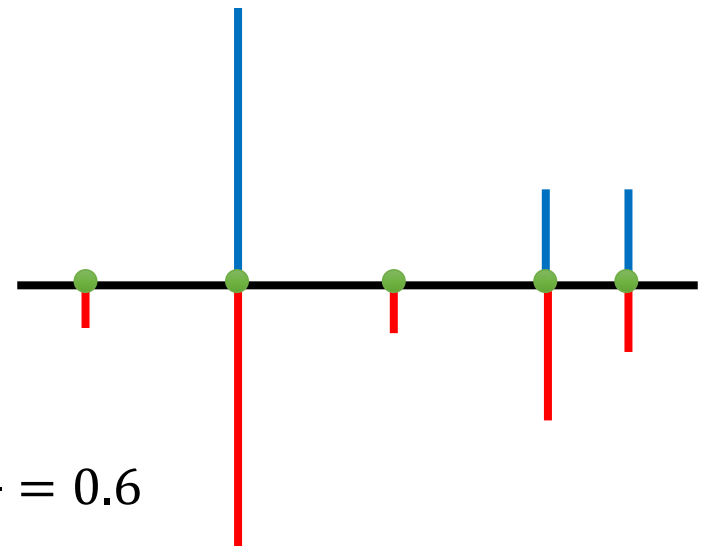


Lib: $\mathbf{b} = (100, 0, 1000, 0, 100, 0, 400, 300)$

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

Reverse dot product

Act.: $\mathbf{a}' = (0, 1000, 0, 300, 300)$

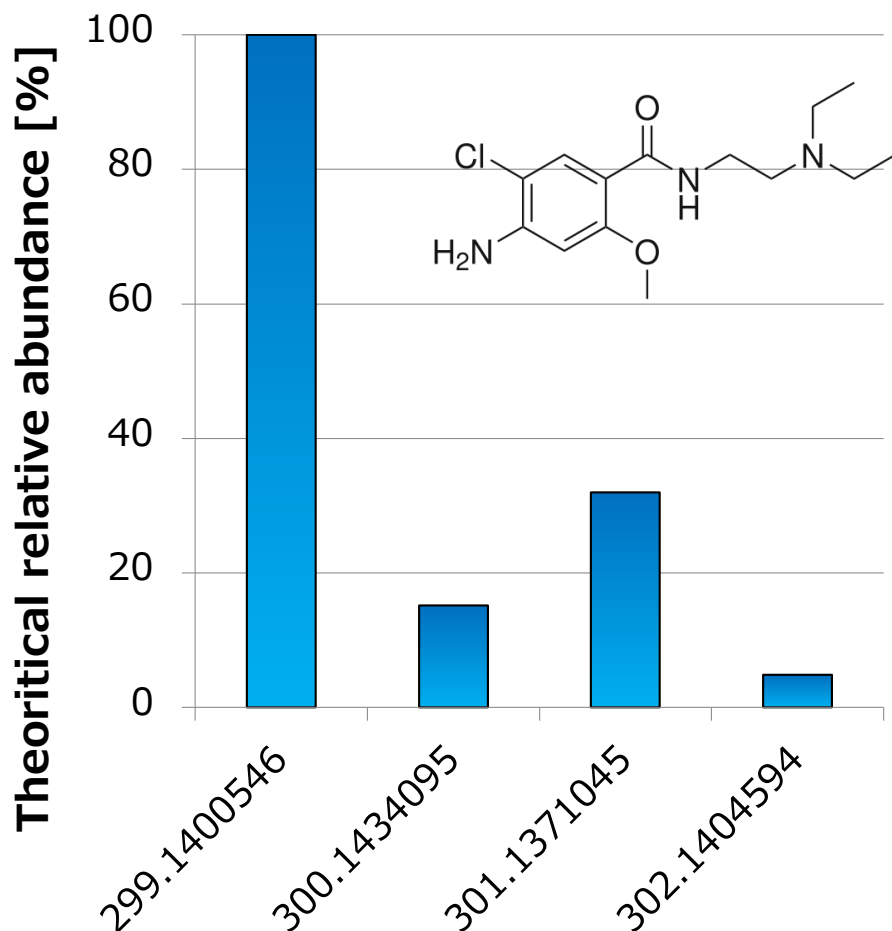


Lib: $\mathbf{b}' = (100, 1000, 100, 400, 300)$

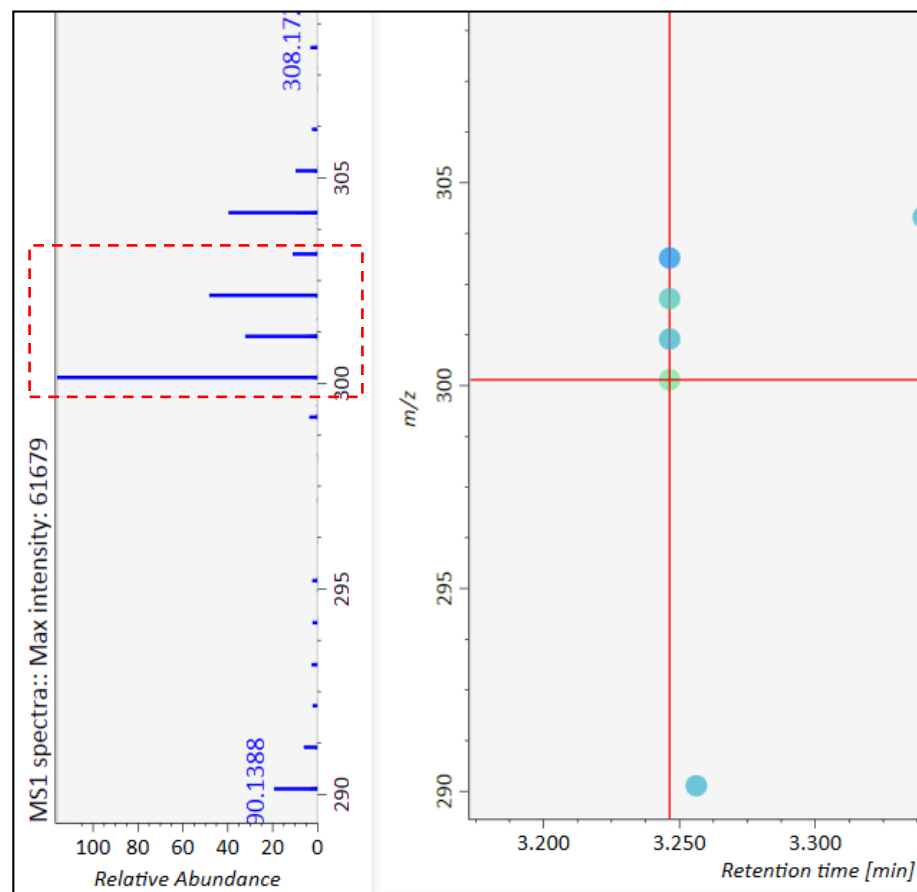
$$\text{Rev. product: } \frac{\mathbf{a}' \cdot \mathbf{b}'}{|\mathbf{a}'| \cdot |\mathbf{b}'|} = 0.995$$

Isotope ratio similarity

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



Similarity: 0.915

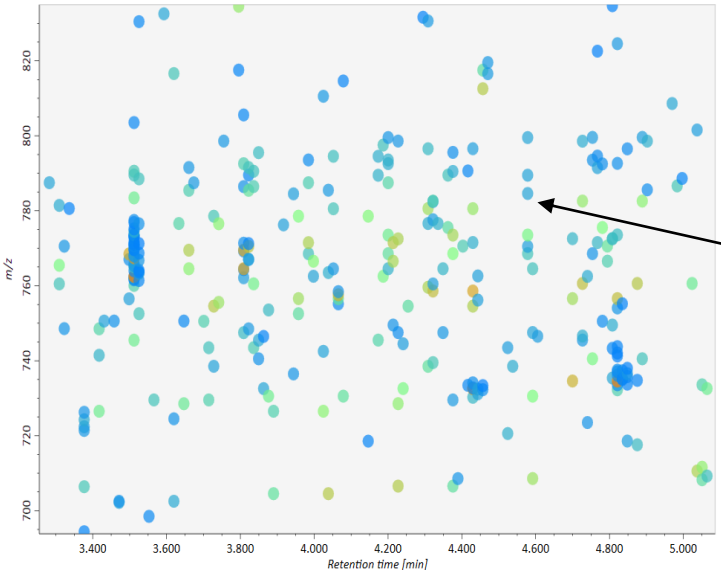


$$S_{ratio} = 1 - \sum |r_{act.i} - r_{lib.i}| \quad r_i = \frac{I_{M+i}}{I_M}, 1 \leq i \leq n$$

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

Analysis parameter setting

Data collection | **Peak detection** | Deconvolution | Identification | Adduct | Alignment

Alignment parameters setting

Result name: alignmentResult_2014_8_31_23

Reference file: Nega_Swath_QC_1_5 ▼

Retention time tolerance: min

MS1 tolerance: Da

Retention time factor: (0-1)

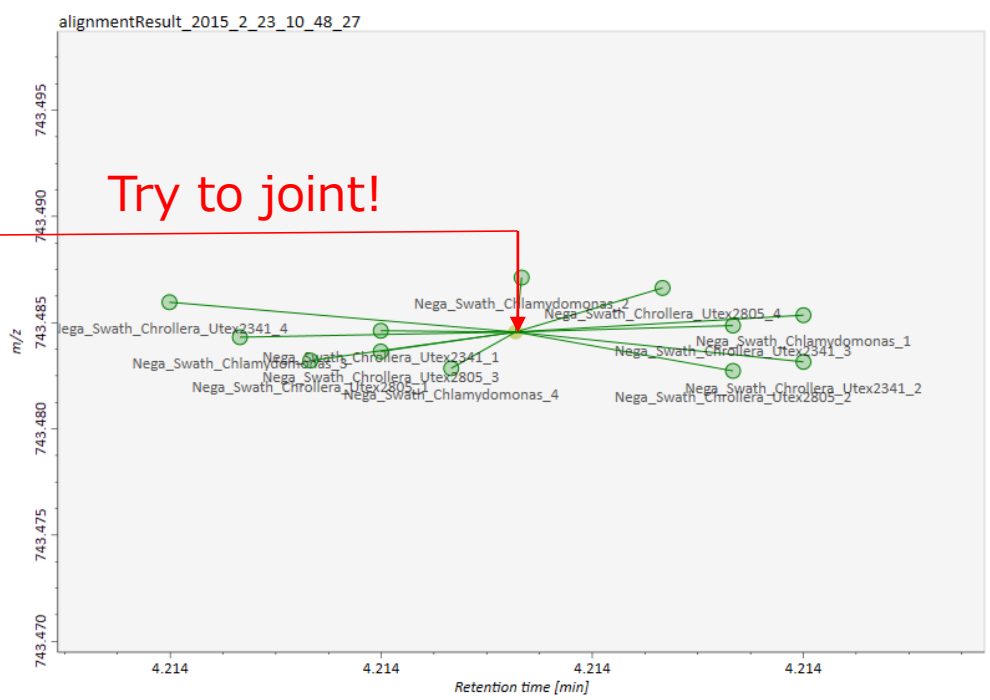
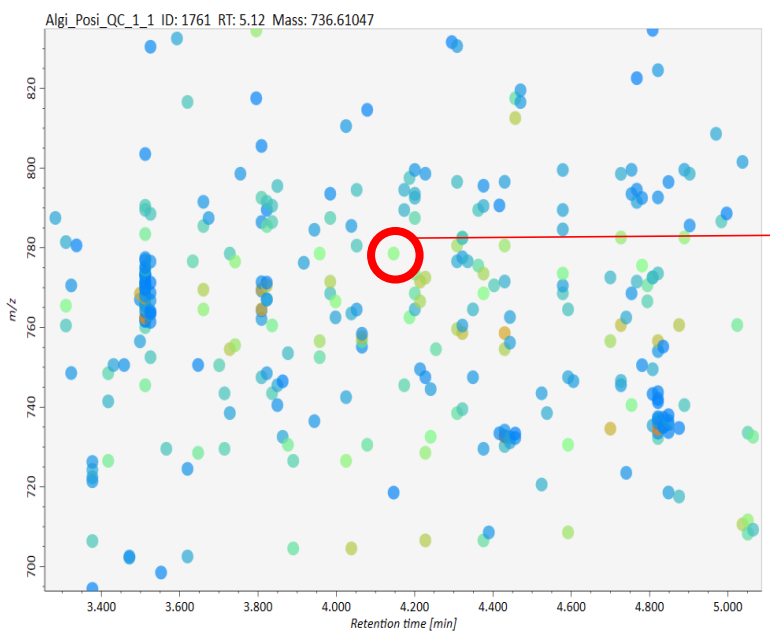
MS1 factor: (0-1)

Peak count filter: %

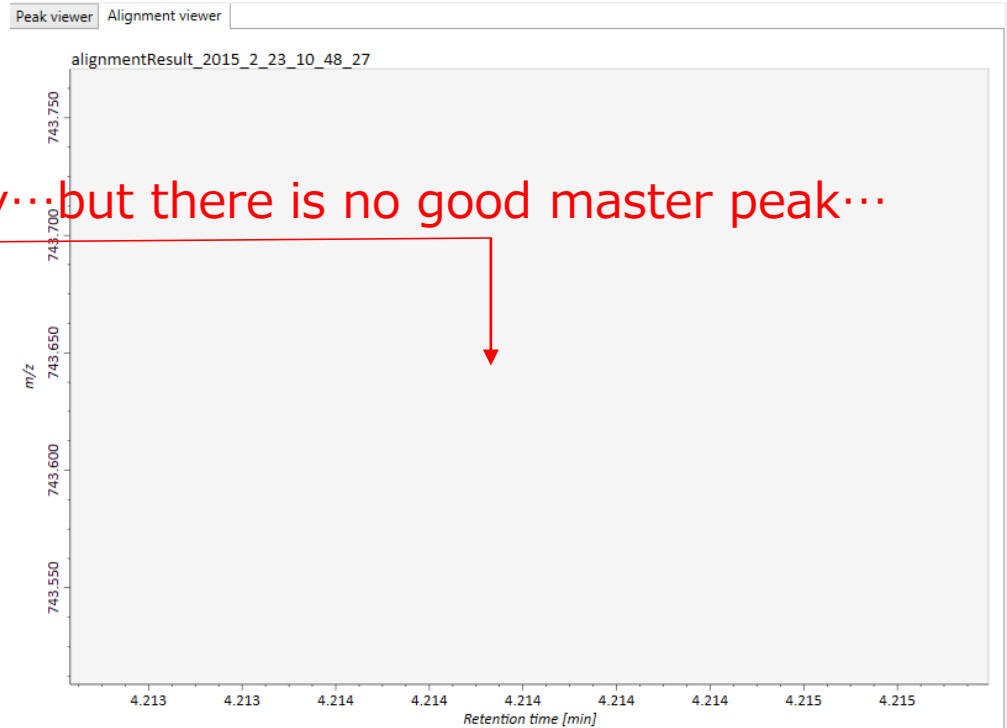
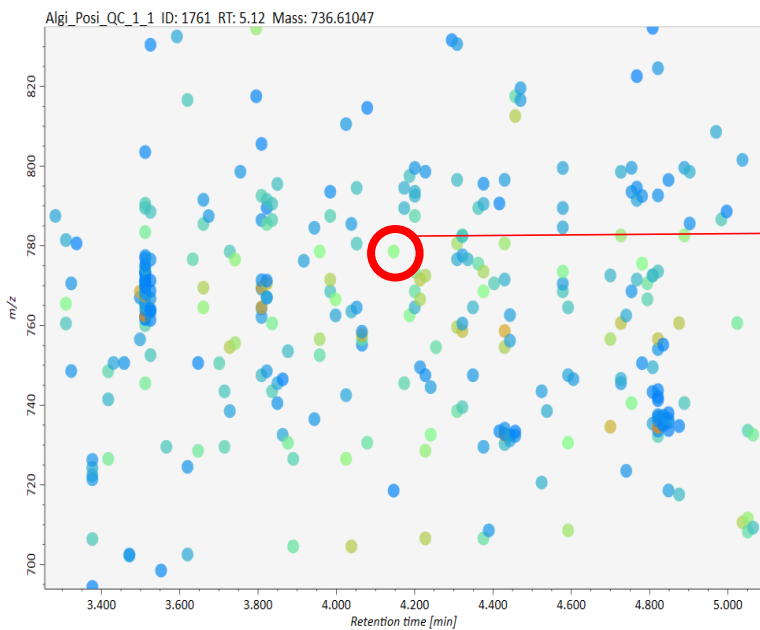
QC at least filter Gap filling option

 Together with Alignment

BTW, what is 'Joint aligner'?



BTW 2, what if there is no master peak?



Making 'master peak list'

a. Making a reference peak table

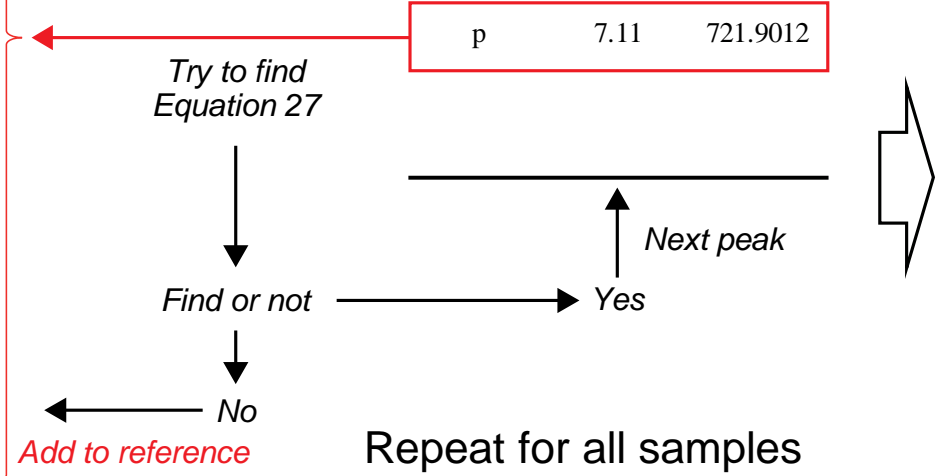
Reference file

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

Sample A

Peak no.	RT [min]	m/z
p	7.11	721.9012

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



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

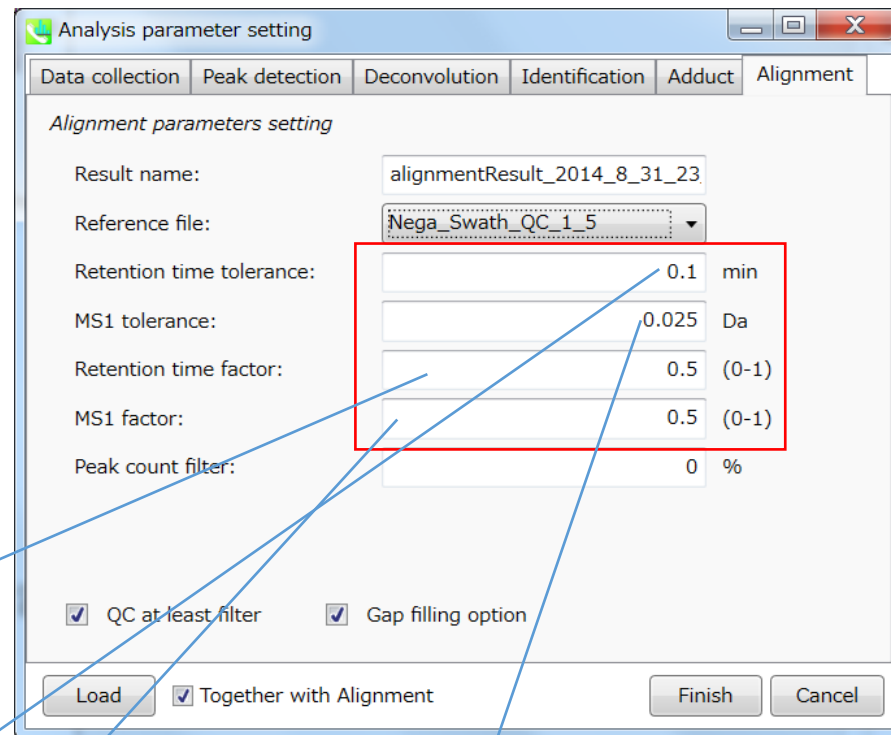
Sample A

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

Joint to matched peak
Equation 28

Repeat for all samples

n-2	11.35	1050.772
n-1	12.85	1001.289
n	12.88	1230.412



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

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

Sample A

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

*Joint to matched peak
Equation 28*

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.



Repeat for all samples

n-2	11.35	1050.772
n-1	12.85	1001.289
n	12.88	1230.412

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

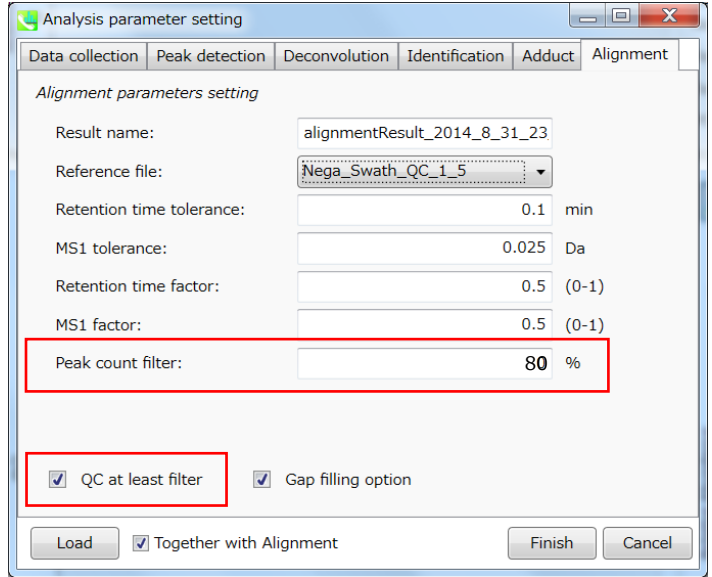
Alignment ID	RT Ave.	m/z Ave.	Sample A	Sample B	Sample C	QC 1	QC 2	QC 3
1	0.72	434.5878	2500	1800	4000	1000	N.D.	2000
2	0.88	541.9050	N.D.	1500	2000	1400	2000	1500
3	1.25	842.1220	53000	62000	40000	45000	30000	35000
4	1.81	254.1079	100	50	730	100	50	730
5	1.99	771.3765	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
6	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)

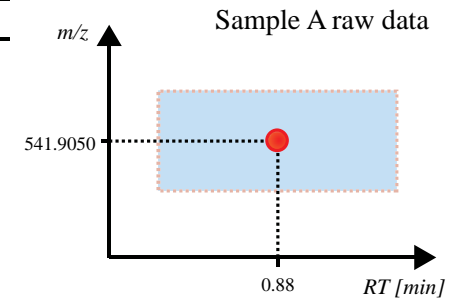


Gap filing: interpolating missing values

d. Interpolating missing values

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

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



Analysis parameter setting

Data collection Peak detection Deconvolution Identification Adduct Alignment

Alignment parameters setting

Result name: alignmentResult_2014_8_31_23

Reference file: Nega_Swath_QC_1_5

Retention time tolerance: 0.1 min

MS1 tolerance: 0.025 Da

Retention time factor: 0.5 (0-1)

MS1 factor: 0.5 (0-1)

Peak count filter: 80 %

QC at least filter Gap filling option

Load Together with Alignment Finish Cancel