
SEQUEST

2020.06

김현우

Introduction

- **Protein**

- An amino acid sequence
 - MEMEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAK**LPKNK**
NRNRYRDVSPFDHSRKLHQEDNDYINASLIKMEEAQRSYIL
TQ....

- **Peptide**

- Substring of the protein
 - **LPKNKNRNRYRDVSPFDHSR**

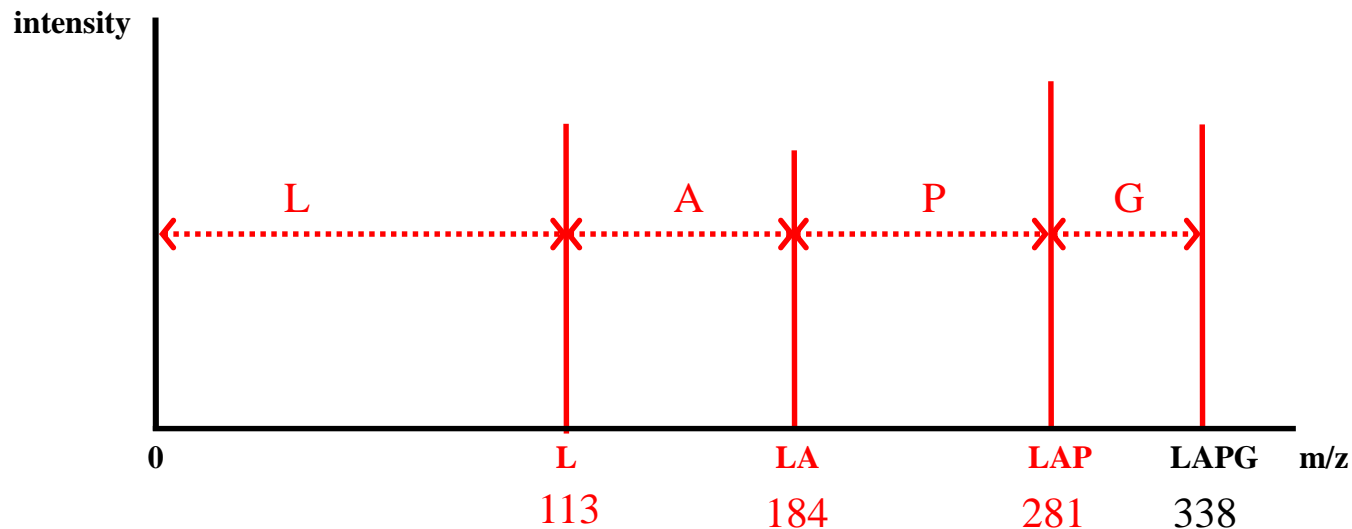
Introduction

- **Tandem mass spectrometry**

- Peptide LAPG

- L, AGP
 - LA, PG
 - LAP, G

L	113
A	71
P	97
G	57



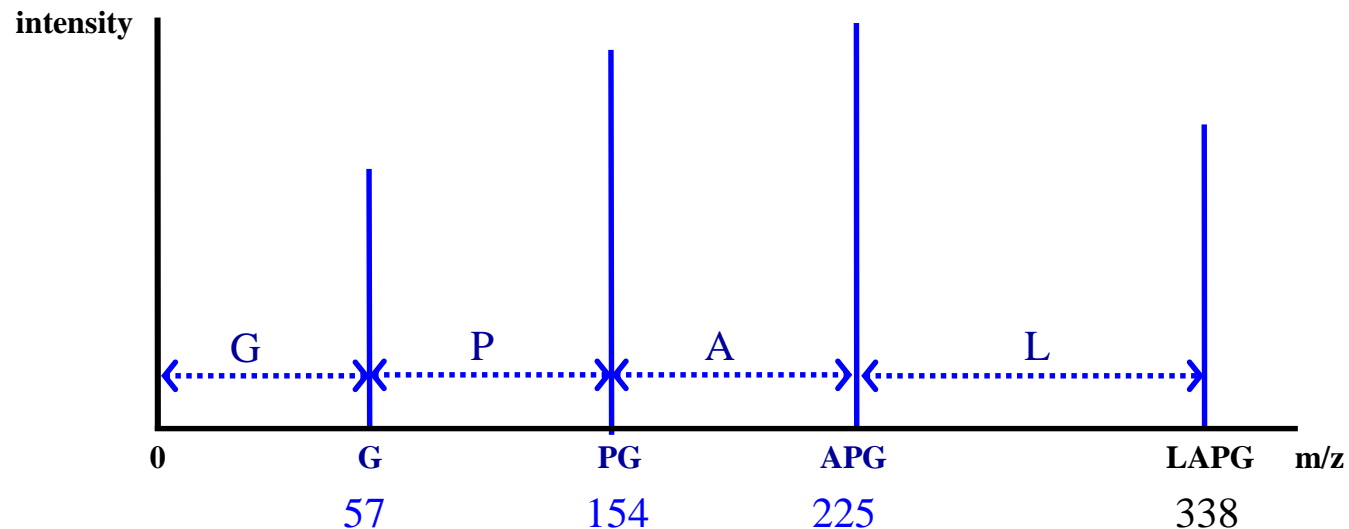
Introduction

- **Tandem mass spectrometry**

- Peptide LAPG

- L, AGP
 - LA, PG
 - LAP, G

L	113
A	71
P	97
G	57



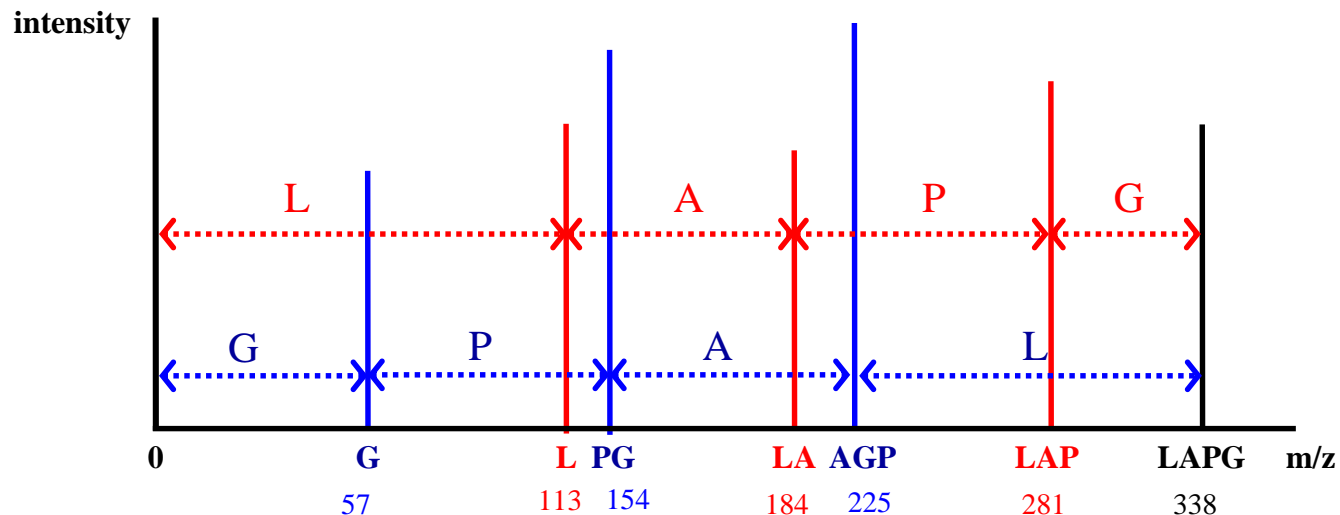
Introduction

- **Tandem mass spectrometry**

- Peptide LAPG

- L, AGP
 - LA, PG
 - LAP, G

L	113
A	71
P	97
G	57



Introduction

- **Tandem mass spectrometry**

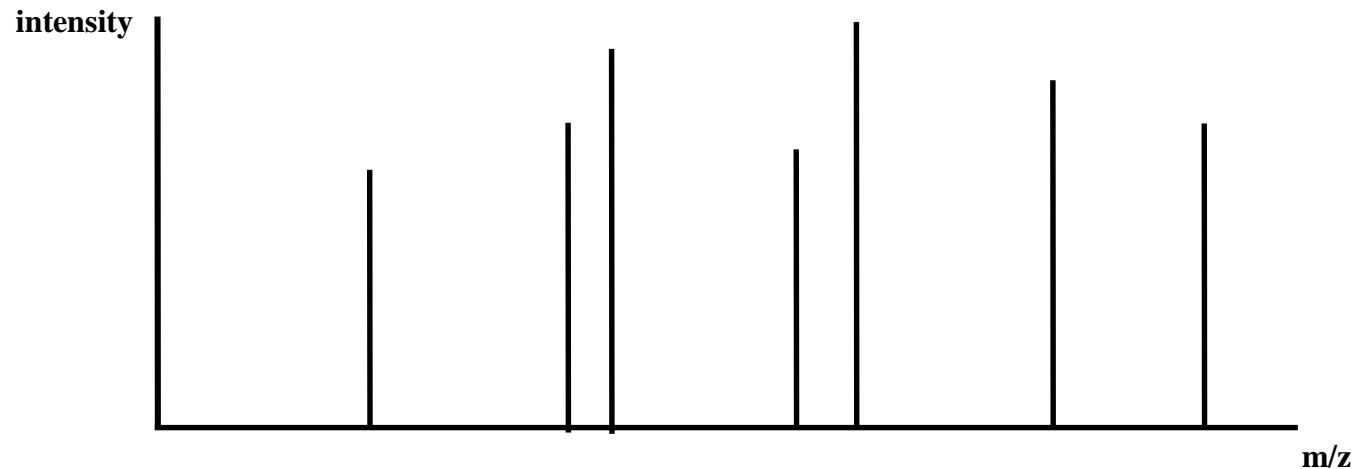
- Peptide LAPG

- L, AGP

- LA, PG

- LAP, G

L	113
A	71
P	97
G	57



Introduction

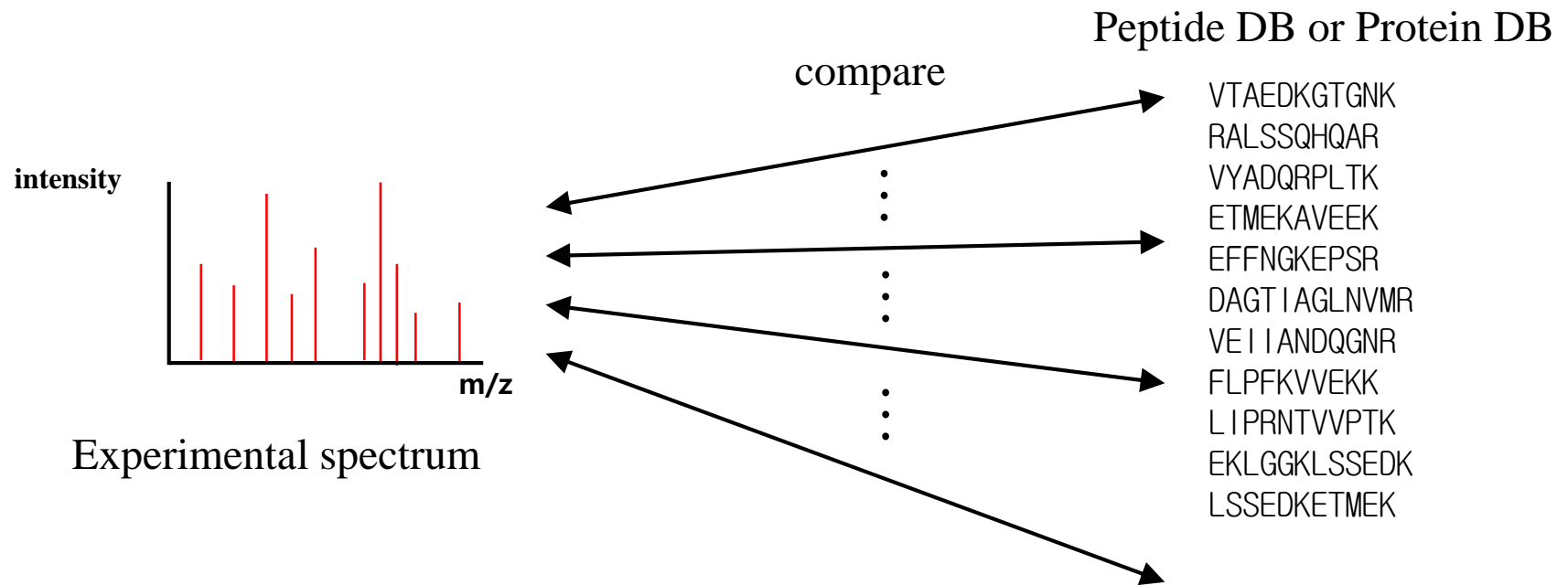
- Peptide identification is an important problem in proteomics
- One of the most popular scoring schemes for peptide identification is XCorr (cross-correlation)
- Since calculating XCorr is computationally intensive, a lot of efforts have been made to develop fast XCorr engines

Introduction

- **Cross-correlation algorithm**
 - Time complexity is $\Theta(n \log n) \rightarrow$ SEQUEST
 - Eng, J.K. *et al.*, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database., *J. am. Soc. Mass Spectrom.*, 1994
 - Time complexity is $\Theta(\tau n) \rightarrow$ Crux
 - Eng, J.K. *et al.*, A Fast SEQUEST Cross Correlation Algorithm., *J. of Proteome Res.*, 2008
 - McIlwain, S. *et al.*, Crux: rapid open source protein tandem mass spectrometry analysis., *J. of Proteome Res.*, 2014
 - Time complexity is $\Theta(\tau + n) \rightarrow$ Tide, Comet
 - Jo, H. *et al.*, Computing Cross-Correlation of SEQUEST in Linear Time., *RECOMB Satellite Conference on Computational Proteomics*, 2010
 - Diamant, B.J. *et al.*, Faster SEQUEST searching for peptide identification from tandem mass spectra., *J. of Proteome Res.*, 2011
 - Eng, J.K. *et al.*, Comet: an open-source MS/MS sequence database search tool., *Proteomics*, 2013

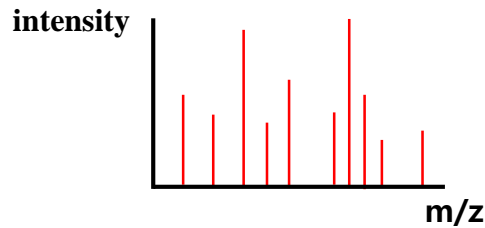
Introduction

- **Peptide identification**



Peptide identification tools

- **Basic algorithm for peptide identification**



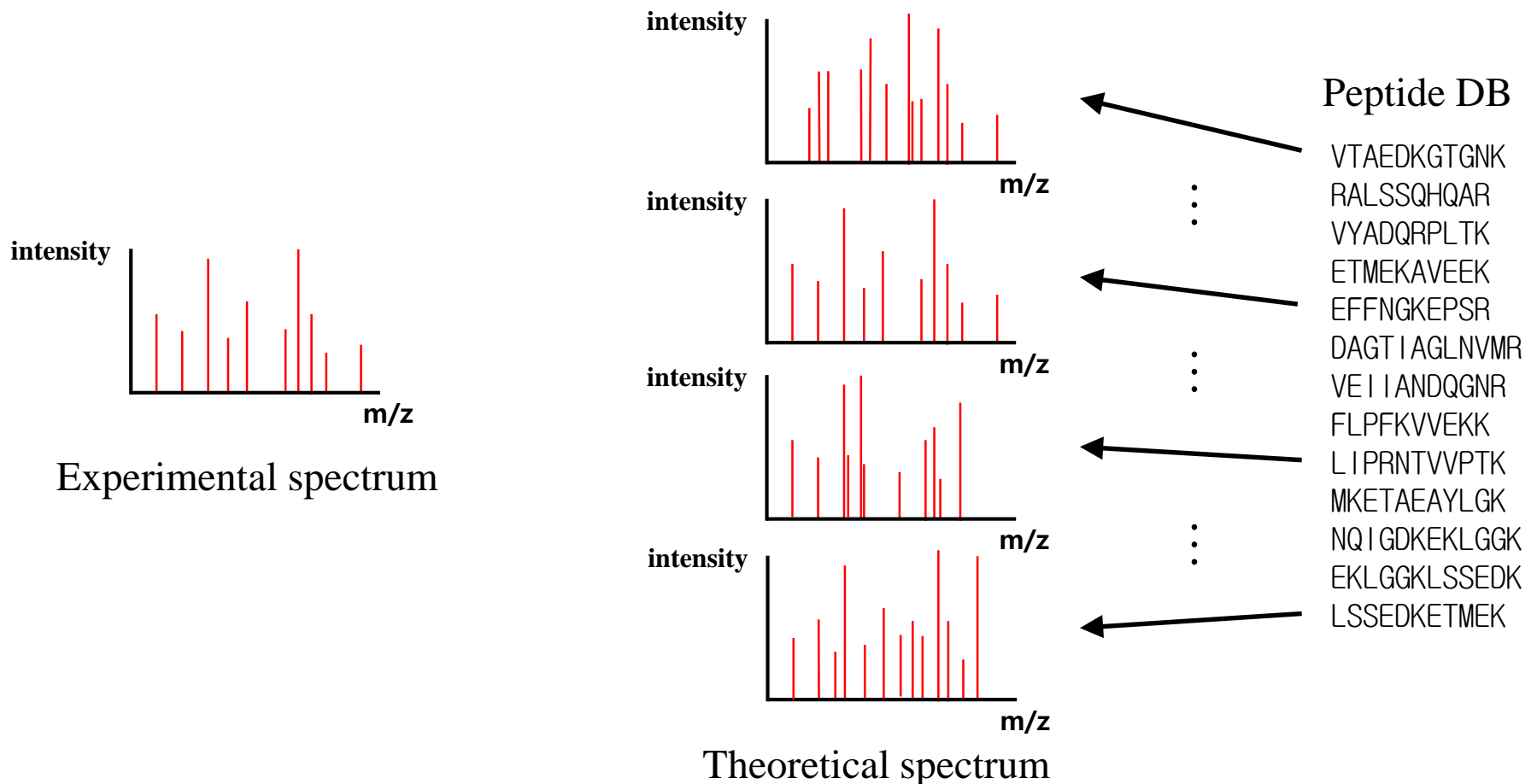
Experimental spectrum

Peptide DB

VTAEDKGTGNK
RALSSQHQAR
VYADQRPLTK
ETMEKAVEEK
EFFNGKEPSR
DAGT I AGLNVMR
VE I I ANDQQNR
FLPFKVVEKK
L I PRNTVVPTK
MKETA EAYLGK
NQ I GDKEKLGGK
EKLGGKLSSDK
LSSDKETMEK

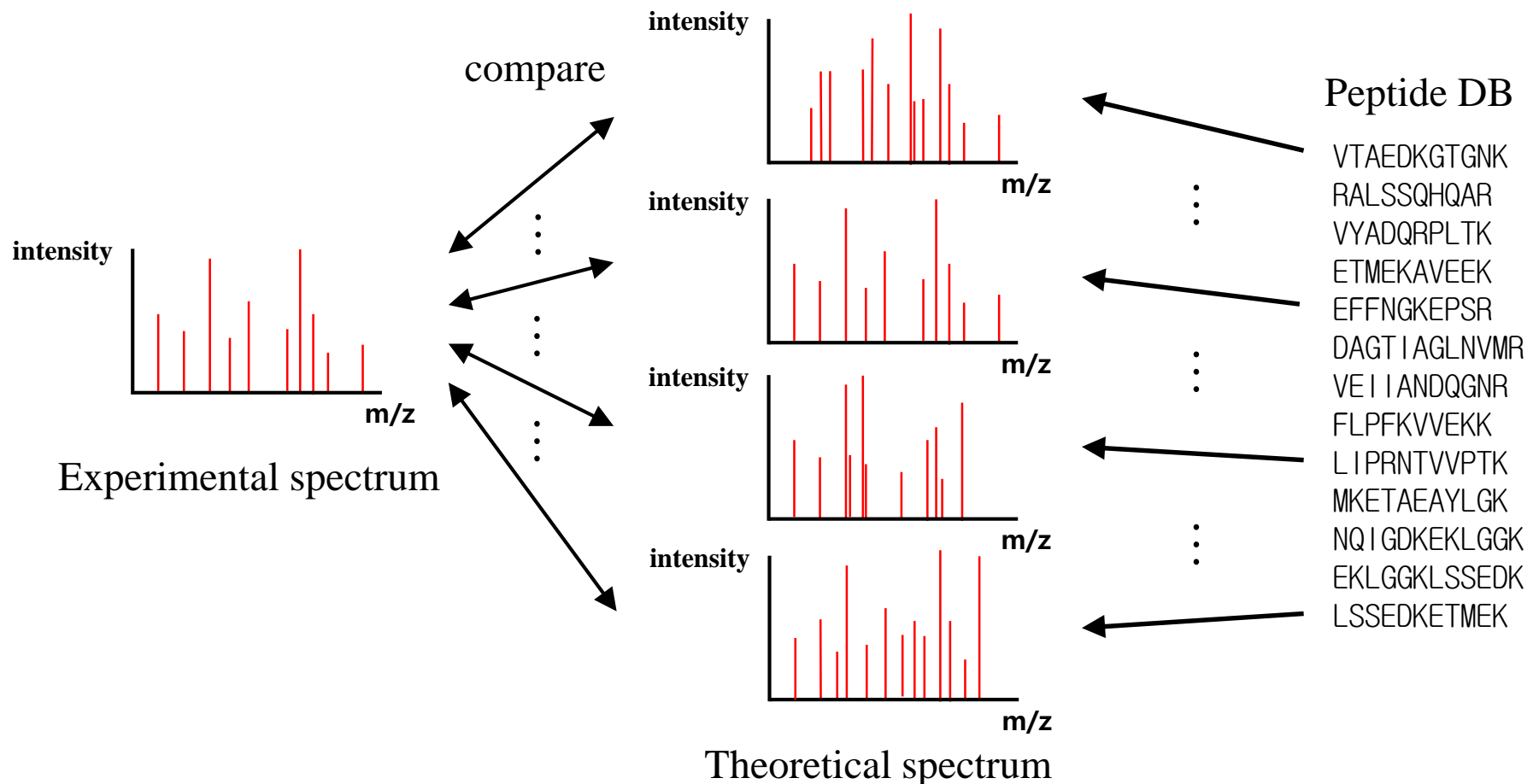
Peptide identification tools

- Basic algorithm for peptide identification



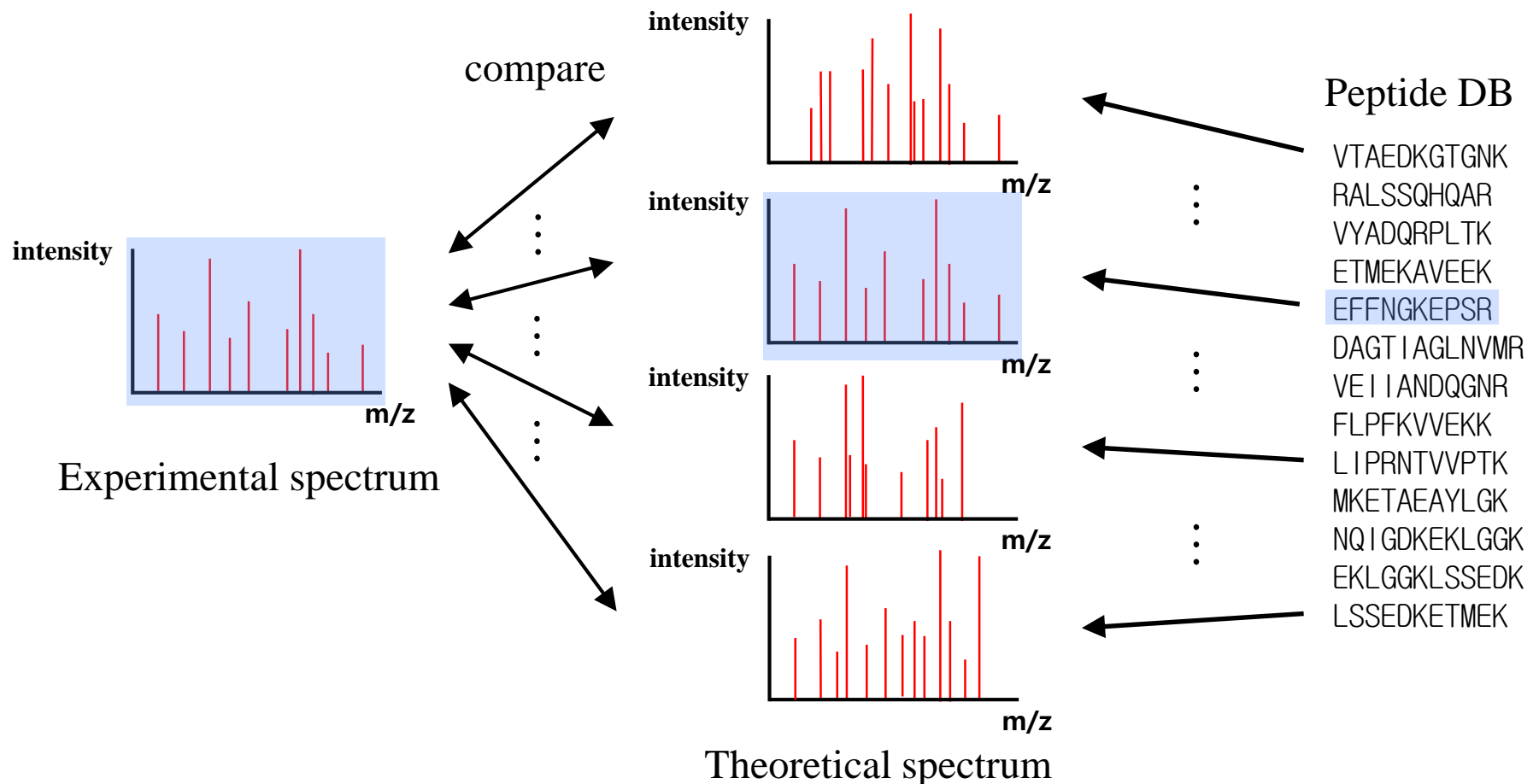
Peptide identification tools

- **Basic algorithm for peptide identification**



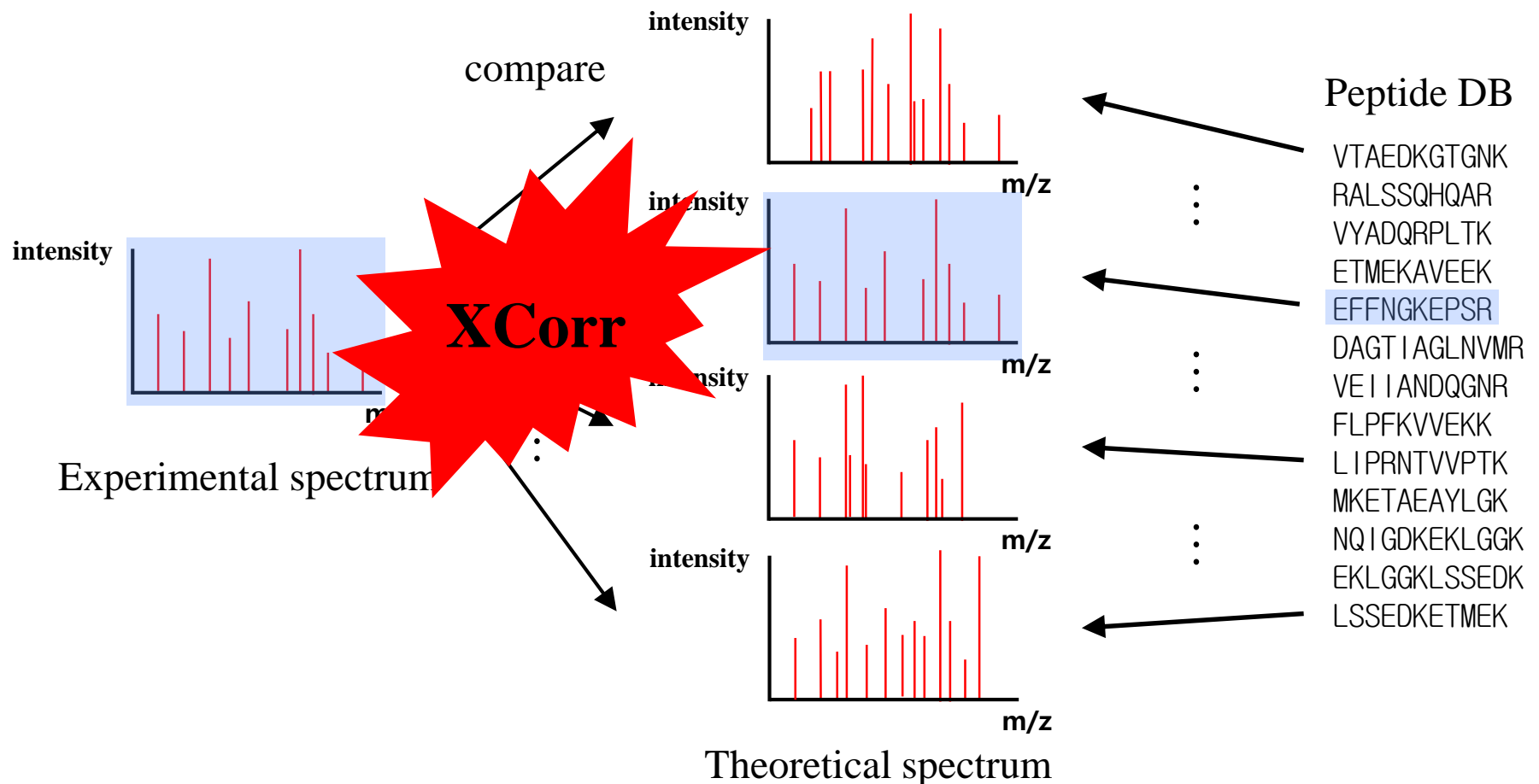
Peptide identification tools

- Basic algorithm for peptide identification



Peptide identification tools

- Basic algorithm for peptide identification



SEQUEST

- **SEQUEST**

1. Tandem mass spectrometry data reduction
2. Database search
3. Filtering by a simple scoring method
4. Cross-correlation

Tandem mass spectrometry data reduction

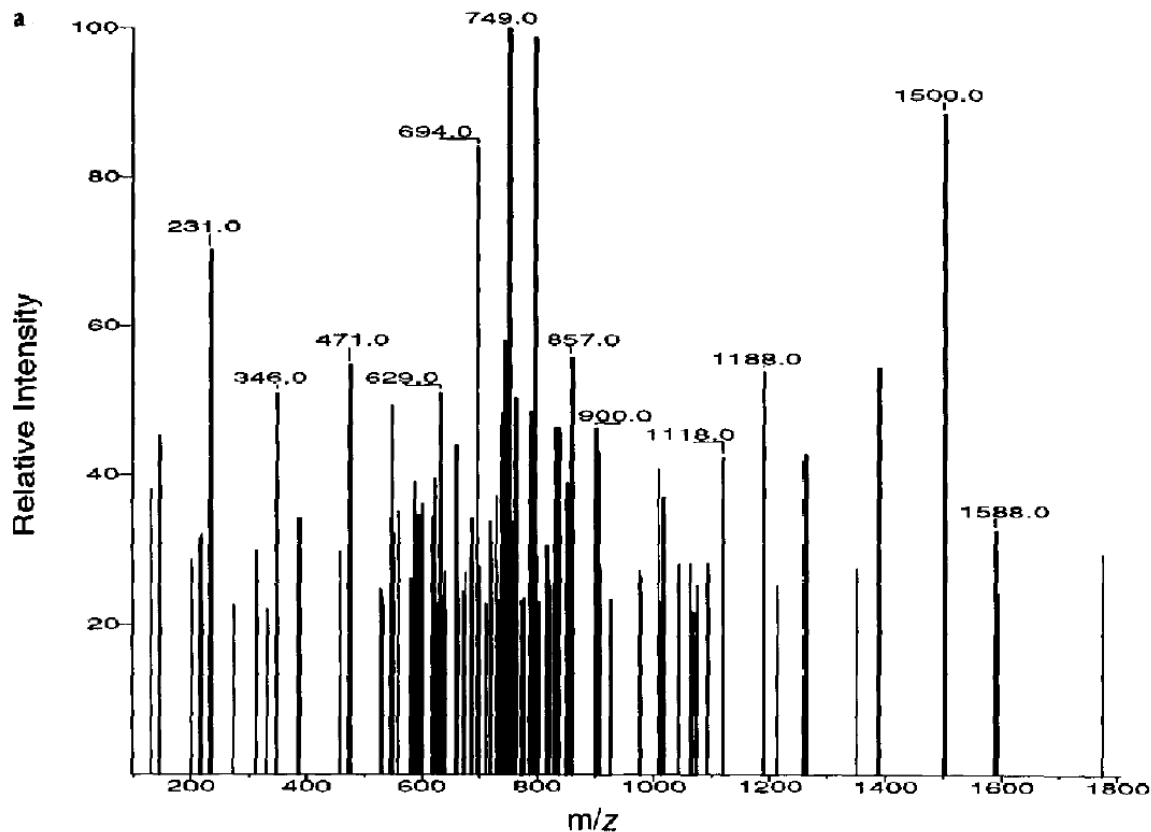
- Fragment ion mass-to-charge ratios are converted to the rounded nominal values(nearest integer).
 - We realize an increase in computational speed.
- A 10-u window around the precursor ion is removed.
 - To eliminate possible matches of a predicted fragment ion to the mass-to charge ratio of the precursor ion.

Tandem mass spectrometry data reduction

- To eliminate noise from the spectrum and to reduce the number of ions to be considered.
 - All but the 200 most abundant ions are removed
 - The remaining ions are renormalized to 100.
 - The abundances of fragment ions within ± 1 u of each other are equalized to the higher value.
- Immonium ions are frequently observed when the amino acids His, Met, Trp, Tyr, or Phe are present in the spectrum.
 - The presence of these ions in the mass spectrum is noted during this preprocessing step.

Tandem mass spectrometry data reduction

- Data for the peptide DLRSWTAADTAAQISQ obtained from the ion $[M + 2H]^{+2}$ at m/z 868.



Database search

- Select each peptide from DB whose mass is within a specified mass tolerance.
- Mass tolerance: $\pm 0.05\%$ or a minimum of 1 u
- Chemical modifications to amino acids can be considered by changing the amino acid masses.

Filtering by a simple scoring method

- Select 500 high scoring peptides.
- $S_P = (\sum i_m)n_i(1 + \beta)(1 + \rho) / n_t$
 - $\sum i_m$: the sum of matched intensities
 - n_i : the number of matched peaks
 - β : continuity of an ion series
 - ρ : immonium ion peaks
 - n_t : the total number of amino acids

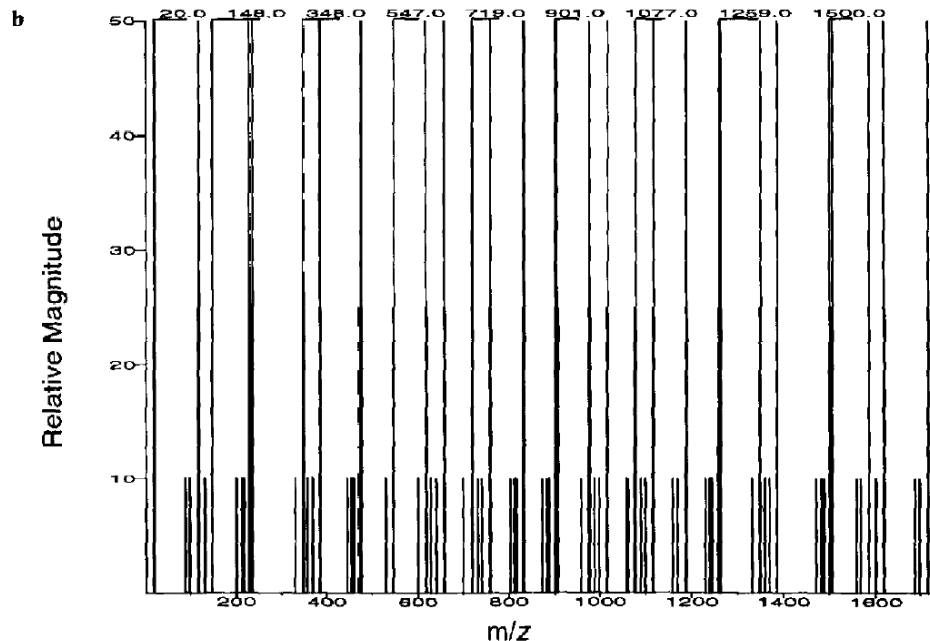
Cross-correlation

1. Generation of theoretical spectrum for every peptide.
2. Preprocessing the experimental spectrum.
3. Compute the cross correlation score between every pair of theoretical spectrum and the experimental spectrum.
4. Select the most promising peptide using the score.

Cross-correlation

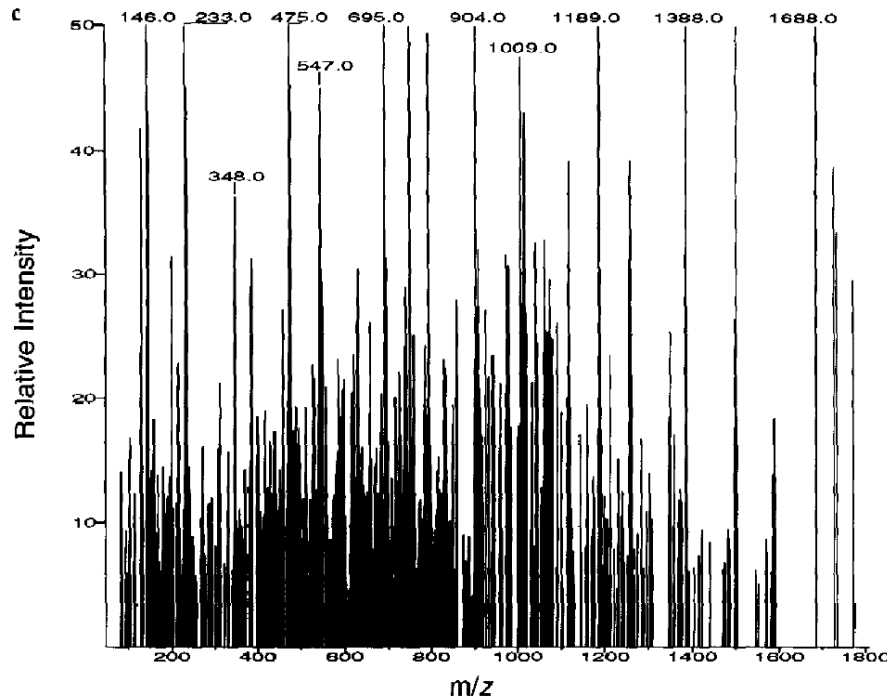
1. Generation of theoretical spectrum for every peptide.

- 50 : b-ions, y-ions
- 25 : within ± 1 of b-ions and y-ions
- 10 : neutral loss of ammonia, water, and carbon monoxide and their ± 1



Cross-correlation

2. Preprocessing the experimental spectrum.
 - Remove the mass-to-charge ratio for the precursor ion.
 - Dividing the original spectrum into 10 equal regions and normalizing the ions in each region to a value of 50.



Cross-correlation

3. Compute the cross-correlation score

- Compute a scoring function.

$$R_{\tau} = \sum_{i=0}^{n-1} x[i]y[i + \tau]$$

- A discrete correlation of two real functions x and y is one member of the discrete Fourier transform pair

$$R_{\tau} \leftrightarrow X_{\tau} Y_{\tau}^*$$

where X and Y are the discrete Fourier transforms of x and y the Y^* denotes complex conjugation.

Cross-correlation

3. Compute the cross-correlation score

- The score
= (The value at $\tau = 0$) –
(the mean of the cross-correlation over $-75 < \tau < 75$)

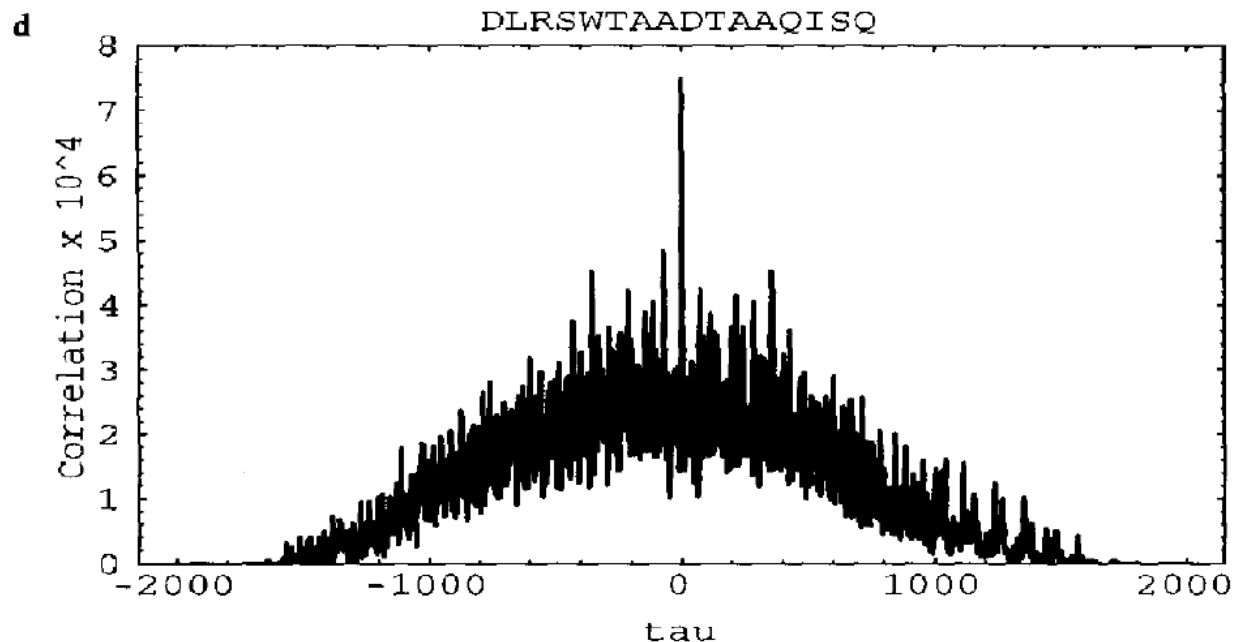


Figure 2. (Continued)

Cross-correlation

4. Select the most promising peptide using the score.
 - The scores are normalized to 1.0 and termed C_n .
 - If $[(C_n \text{ for the highest}) - (C_n \text{ for the second highest})]$ is larger than $0.1 (\Delta C_n > 0.1)$, it is usually correct.

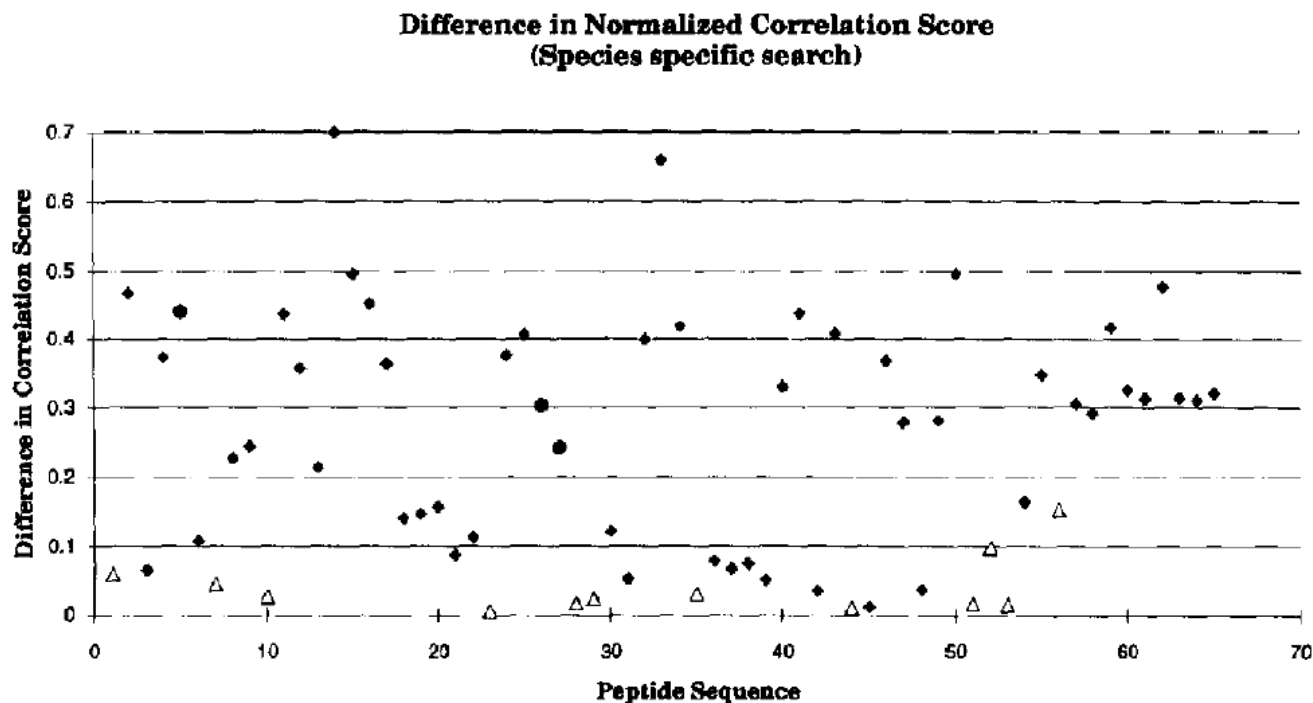
Cross-correlation

4. Select the most promising peptide using the score.

No.	Db	Protein identified	Sequence identified	S ΔC_n	S C_n	G C_n
42	E	trypsin, bovine	SSGTSYPDVLK	0.035	1	1
43	E	trypsin, bovine	TLNNDIMLIK	0.407	1	1
44	E	trypsin, bovine	SIVHPSYNSNTLNNDIMLIK	0.012	4	17
45	E	trypsin, bovine	VASISLPTSCASAGTQCLISGWGNTK	0.012	1	7
46	E	trypsin, bovine	VASISLPTSCASAGTQCLISGWGNTK	0.367	1	1
47	E	tufA gene product, <i>E. coli</i>	VGEEVEIVGIK	0.278	1	1
48	E	tufA gene product, <i>E. coli</i>	VTLIHPIAMDDGLR	0.037	1	2
49	E	ORF-D, <i>E. coli</i>	GGDTVTLNEDLTQIPK	0.281	1	1
50	E	GAD α protein, <i>E. coli</i>	GWQVPAFTLGGEATDIVMR	0.494	1	1
51	H	mRNA for HLA-DR antigens	MATPLLMQALP	0.017	5	—
52	H	mRNA for HLA-DR antigens	MATPLLMQALPM	0.098	2	9
53	H	mRNA for HLA-DR antigens	KPPKPVSKMR*	0.015	7	26
54	H	mRNA for HLA-DR antigens	PKPPKPVSKMR*	0.163	1	2
55	H	lymphocyte antigen	DLRSWTAADTAAQISQ	0.347	1	1
56	H	MHC Class I HLA	DLRSWTAADTAAQITQ	0.152	2	2
57	Y	enolase gene, <i>Sc</i>	EEALDLIVDAIK	0.305	1	1
58	Y	enolase gene, <i>Sc</i>	NPTVEVELTTEK	0.292	1	1
59	Y	enolase gene, <i>Sc</i>	DPFAEDDWEAWSH	0.416	1	1
60	Y	pyruvate kinase, <i>Sc</i>	LPGTDVDLPALSEK	0.325	1	1
61	Y	hexokinase PI gene, <i>Sc</i>	IEDDPFVFLEDTDIFQK	0.312	1	1
62	Y	hypusine protein HP2, <i>Sc</i>	APEGELGDSLQTADEGK	0.476	1	1
63	Y	chromosome III compl. DNA, <i>Sc</i>	IPAGWQGLDNGPESR	0.314	1	1
64	Y	chromosome III compl. DNA, <i>Sc</i>	TGGGASLELLEGG	0.310	1	1
65	Y	BMH1 gene product, <i>Sc</i>	QAFDDAIAELDTLSEESYK	0.321	1	1

Cross-correlation

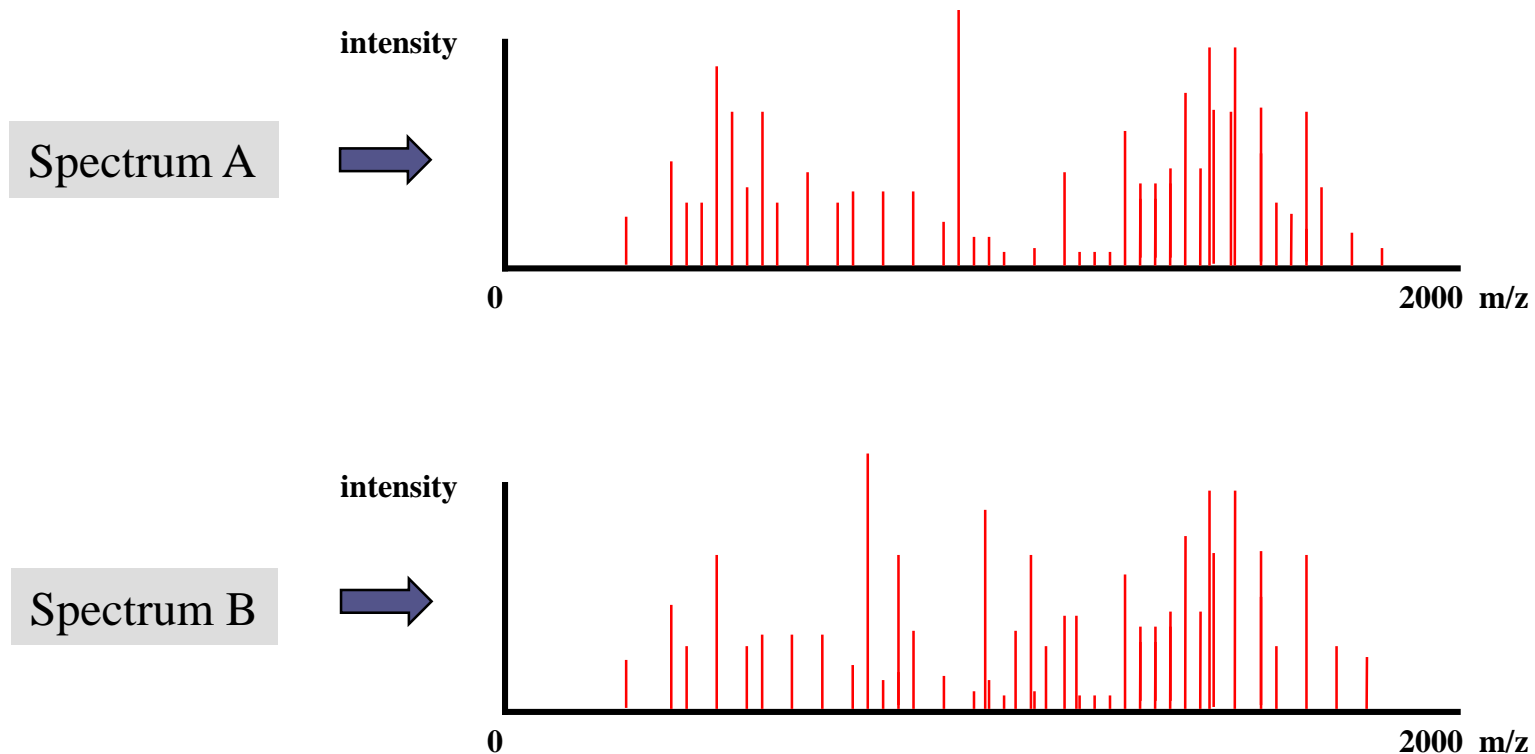
4. Select the most promising peptide using the score.



- ♦ Top score is correct
- Δ Top score is incorrect
- * The correct sequence is within the top two answers which have very close scores and differ by only one amino acid. The displayed ΔC_n is between 1st and 3rd scores.

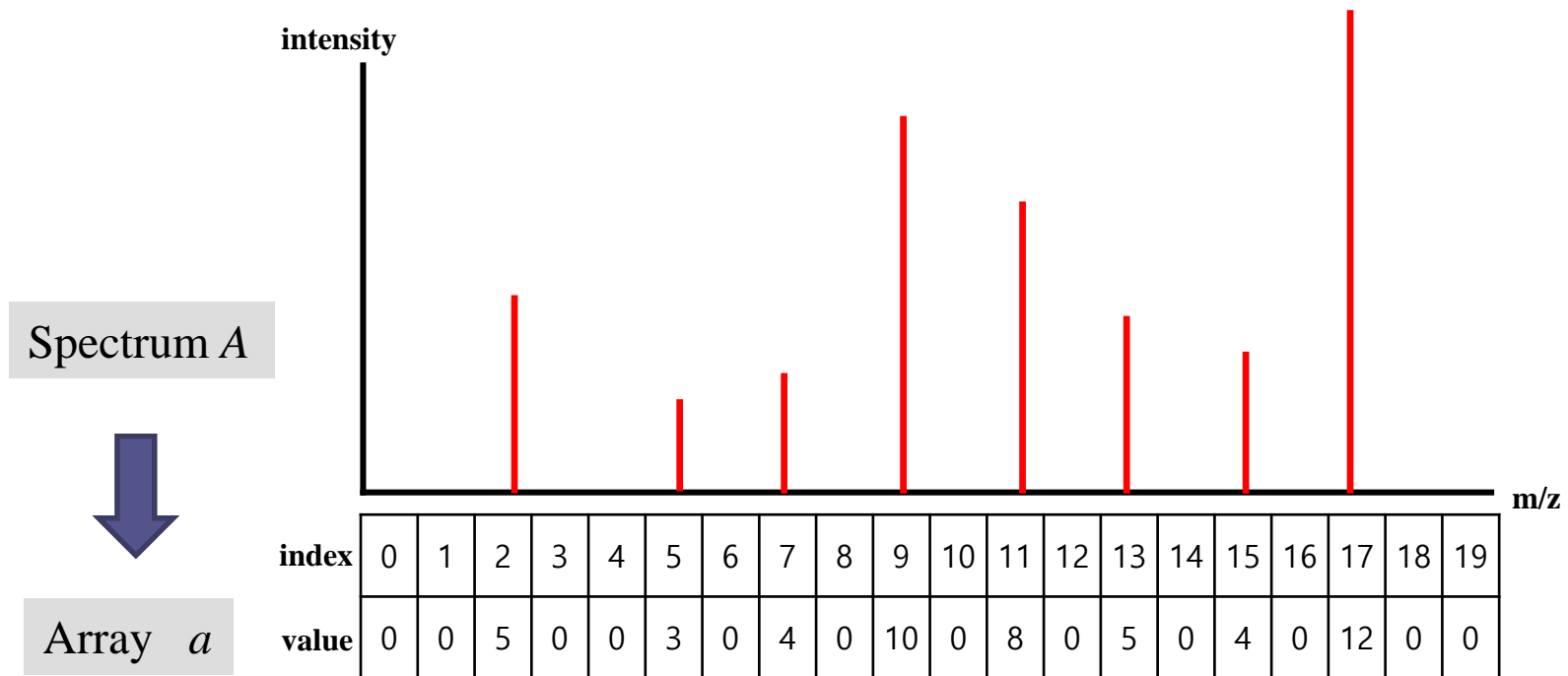
Cross-correlation

- **Cross-correlation**
 - Measures the similarity of two spectra



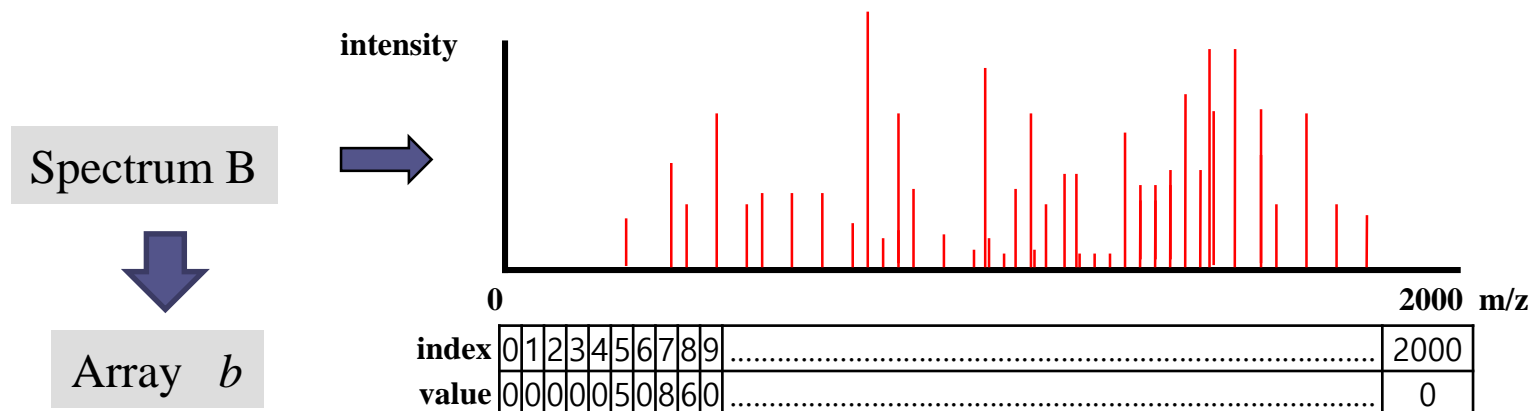
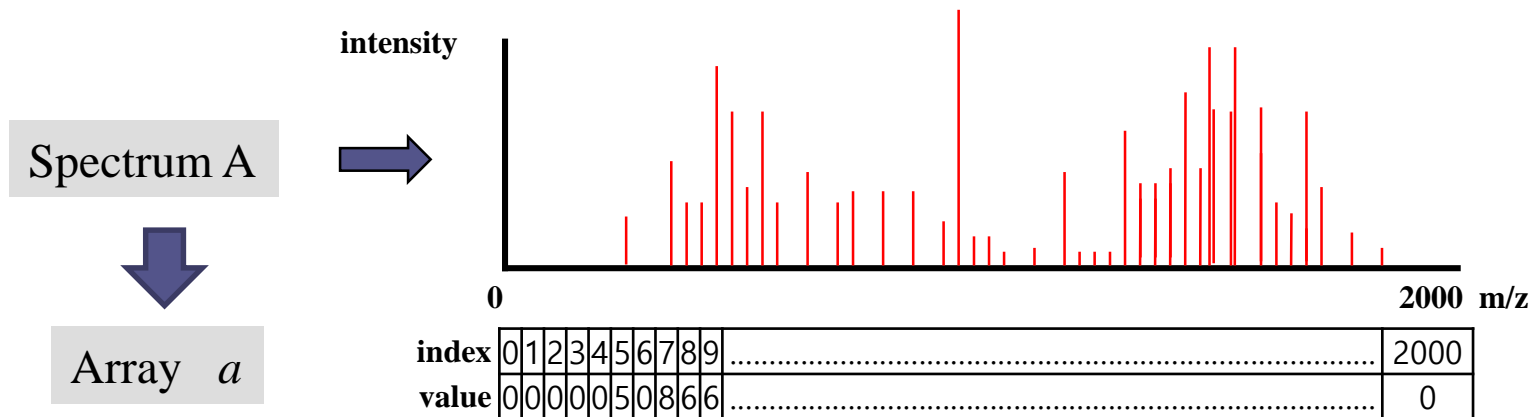
Cross-correlation

- Cross-correlation**



Cross-correlation

- **Cross-correlation**



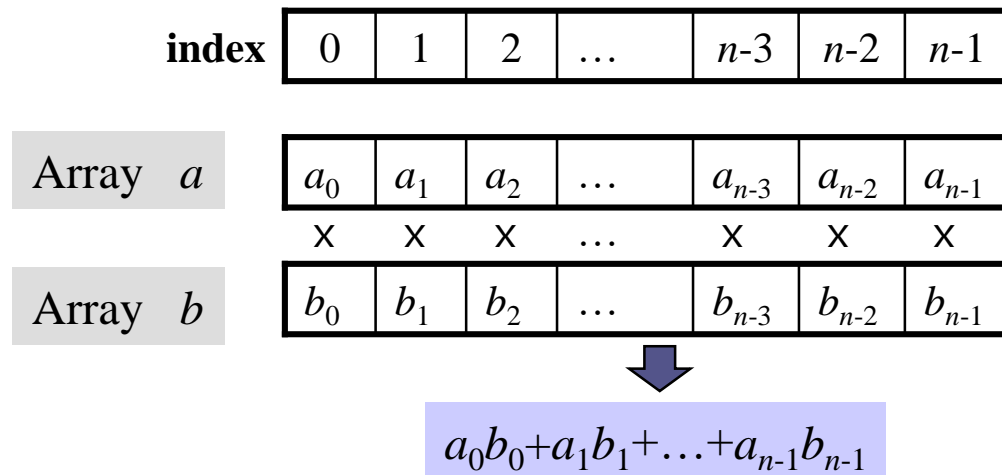
Cross-correlation

- **Cross-correlation for peptide identification**
 - XCorr Score = XCorr value at $\tau = 0$
 - Mean of XCorr values over $-75 < \tau < 75$

Cross-correlation

- Cross-correlation for peptide identification

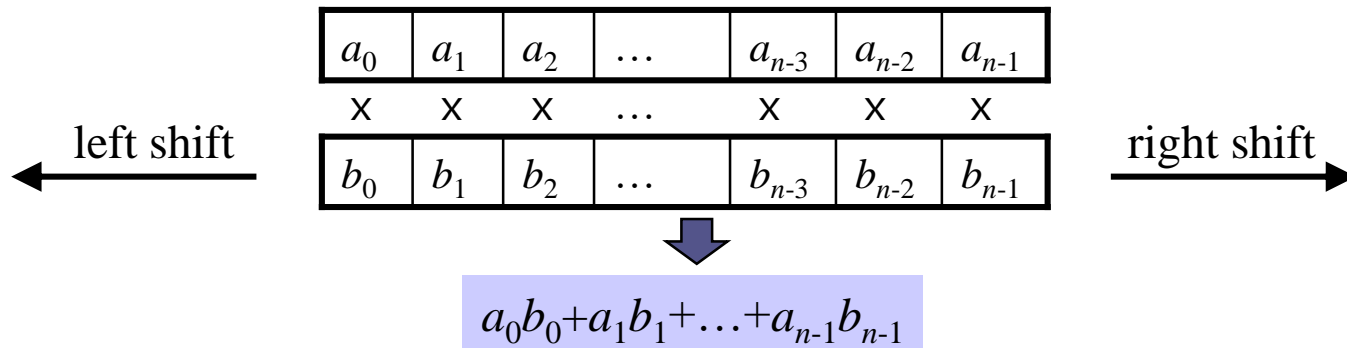
$$\tau = 0$$



Cross-correlation

- **Cross-correlation for peptide identification**

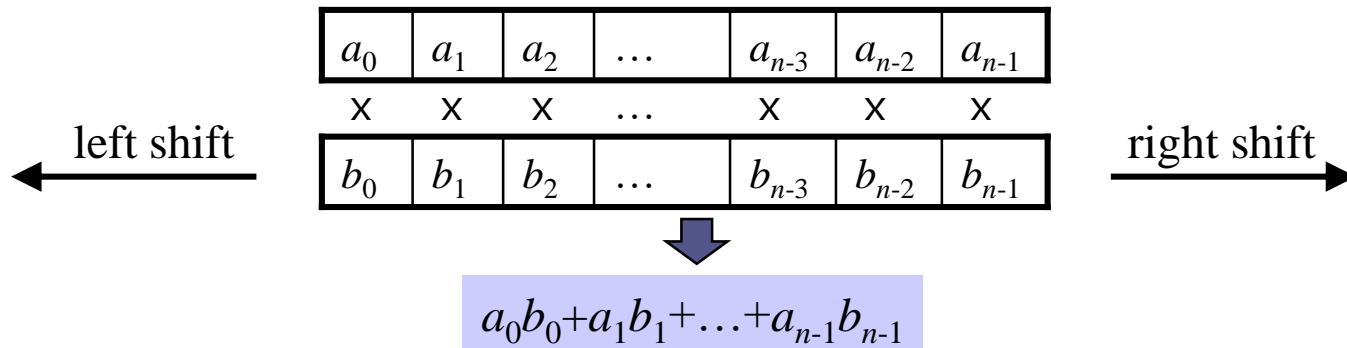
$$\tau = 0$$



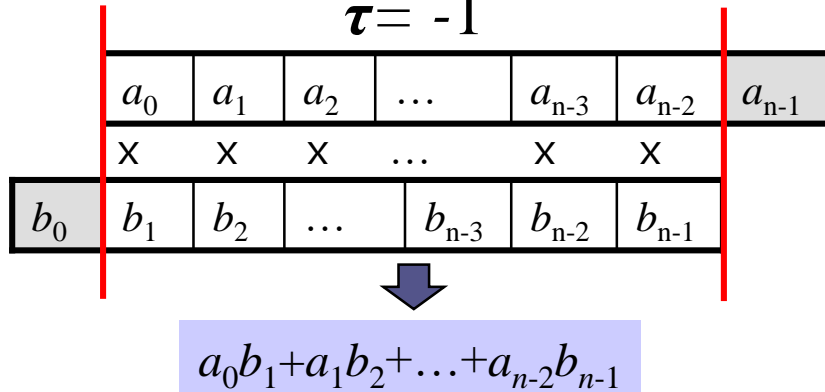
Cross-correlation

- Cross-correlation for peptide identification**

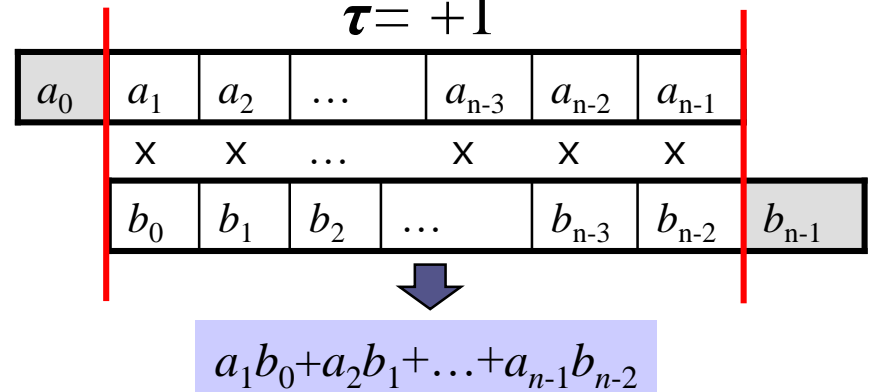
$$\tau = 0$$



$$\tau = -1$$



$$\tau = +1$$



Cross-correlation

- **Cross-correlation for peptide identification**
 - XCorr Score = XCorr value at $\tau = 0$
 - Mean of XCorr values over $-75 < \tau < 75$

Cross-correlation

- How to compute mean of the XCorr values over $-75 < \tau < 75$



$$\frac{1}{149} \sum_{\tau=-74}^{74} \sum_{i=0}^{n-1} a[i]b[i+\tau] = \frac{1}{149} (a_0b_{74} + a_1b_{75} + \dots + a_{n-75}b_{n-1} \\ + a_0b_{73} + a_1b_{74} + \dots + a_{n-74}b_{n-1} \\ \vdots \\ + a_0b_0 + a_1b_1 + \dots + a_{n-1}b_{n-1} \\ \vdots \\ + a_{73}b_0 + a_{74}b_1 + \dots + a_{n-1}b_{n-74} \\ + a_{74}b_0 + a_{75}b_1 + \dots + a_{n-1}b_{n-75})$$

Linear time algorithm

- Time complexity is $\Theta(\tau n)$

$$\begin{aligned}
 \sum_{\tau=-74}^{74} \sum_{i=0}^{n-1} a[i]b[i+\tau] &= a_0b_{74} + a_1b_{75} + \cdots + a_{n-75}b_{n-1} \quad \text{||} \longrightarrow \tau=-74 \\
 &+ a_0b_{73} + a_1b_{74} + \cdots + a_{n-74}b_{n-1} \quad \text{||} \longrightarrow \tau=-73 \\
 &\vdots \\
 &+ a_0b_0 + a_1b_1 + \cdots + a_{n-1}b_{n-1} \quad \text{||} \longrightarrow \tau=0 \\
 &\vdots \\
 &+ a_{73}b_0 + a_{74}b_1 + \cdots + a_{n-1}b_{n-74} \quad \text{||} \longrightarrow \tau=73 \\
 &+ a_{74}b_0 + a_{75}b_1 + \cdots + a_{n-1}b_{n-75} \quad \text{||} \longrightarrow \tau=74
 \end{aligned}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

$$\begin{array}{r}
 a_0 (b_0 + \cdots + b_{74}) \\
 + a_1 (b_0 + \cdots + b_{74} + b_{75}) \\
 \vdots \\
 + a_{74} (b_0 + \cdots + b_{147} + b_{148}) \\
 \hline
 + a_{75} (b_1 + \cdots + b_{148} + b_{149}) \\
 \vdots \\
 + a_{n-75} (b_{n-149} + \cdots + b_{n-2} + b_{n-1}) \\
 \hline
 + a_{n-74} (b_{n-148} + b_{n-147} + \cdots + b_{n-1}) \\
 \vdots \\
 + a_{n-2} (b_{n-76} + b_{n-75} + \cdots + b_{n-1}) \\
 + a_{n-1} (b_{n-75} + \cdots + b_{n-1})
 \end{array}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

$$\begin{array}{rcl}
 a_0 (b_0 + \cdots + b_{74}) & & \\
 + a_1 (b_0 + \cdots + b_{74} + b_{75}) & & \\
 \vdots & & \\
 + a_{74} (b_0 + \cdots + b_{147} + b_{148}) & & \\
 \hline
 + a_{75} (b_1 + \cdots + b_{148} + b_{149}) & & \\
 \vdots & & \\
 + a_{n-75} (b_{n-149} + \cdots + b_{n-2} + b_{n-1}) & & \\
 \hline
 + a_{n-74} (b_{n-148} + b_{n-147} + \cdots + b_{n-1}) & & \\
 \vdots & & \\
 + a_{n-2} (b_{n-76} + b_{n-75} + \cdots + b_{n-1}) & & \\
 + a_{n-1} (b_{n-75} + \cdots + b_{n-1}) & &
 \end{array}
 \begin{array}{l}
 \left. \begin{array}{l} \\ \\ \\ \\ \end{array} \right\} \text{add} \\
 \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{add+subtract} \\
 \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{subtract}
 \end{array}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

$$\begin{array}{rcl}
 a_0 (b_0 + \cdots + b_{74}) & \longrightarrow \Theta(\tau) & \\
 + a_1 (b_0 + \cdots + b_{74} + b_{75}) & \longrightarrow \Theta(1) & \\
 \vdots & & \\
 + a_{74} (b_0 + \cdots + b_{147} + b_{148}) & \longrightarrow \Theta(1) & \\
 \hline
 + a_{75} (b_1 + \cdots + b_{148} + b_{149}) & \longrightarrow \Theta(1) & \\
 \vdots & & \\
 + a_{n-75} (b_{n-149} + \cdots + b_{n-2} + b_{n-1}) & \longrightarrow \Theta(1) & \\
 \hline
 + a_{n-74} (b_{n-148} + b_{n-147} + \cdots + b_{n-1}) & \longrightarrow \Theta(1) & \\
 \vdots & & \\
 + a_{n-2} (b_{n-76} + b_{n-75} + \cdots + b_{n-1}) & \longrightarrow \Theta(1) & \\
 + a_{n-1} (b_{n-75} + \cdots + b_{n-1}) & \longrightarrow \Theta(1) &
 \end{array}
 \begin{array}{l}
 \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \text{add} \\
 \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{add+subtract} \\
 \left. \begin{array}{l} \\ \end{array} \right\} \text{subtract}
 \end{array}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

$$\begin{array}{rcl}
 a_0 (b_0 + \cdots + b_{74}) & \longrightarrow & \Theta(\tau) \\
 + a_1 (b_0 + \cdots + b_{74} + b_{75}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{74} (b_0 + \cdots + b_{147} + b_{148}) & \longrightarrow & \Theta(1) \\
 \hline
 + a_{75} (b_1 + \cdots + b_{148} + b_{149}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{n-75} (b_{n-149} + \cdots + b_{n-2} + b_{n-1}) & \longrightarrow & \Theta(1) \\
 \hline
 + a_{n-74} (b_{n-148} + b_{n-147} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{n-2} (b_{n-76} + b_{n-75} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1) \\
 + a_{n-1} (b_{n-75} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1)
 \end{array}
 \begin{array}{l}
 \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \text{add} \\
 \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \text{add+subtract} \\
 \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{subtract}
 \end{array}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

$$\begin{array}{rcl}
 & \vdots & \\
 + a_{74} (b_0 + b_1 + b_2 + b_3 + & \dots & + b_{148}) \\
 + a_{75} (\text{red box} b_1 + b_2 + b_3 + & \dots & + b_{148} \text{blue box} + b_{149}) \\
 + a_{76} (& \text{red box} & b_2 + b_3 + \dots + b_{148} + b_{149} \text{blue box} + b_{150}) \\
 + a_{77} (& & \text{red box} b_3 + \dots + b_{148} + b_{149} + b_{150} \text{blue box} + b_{151}) \\
 & \vdots &
 \end{array}$$

Linear time algorithm

- Time complexity is $\Theta(\tau + n)$

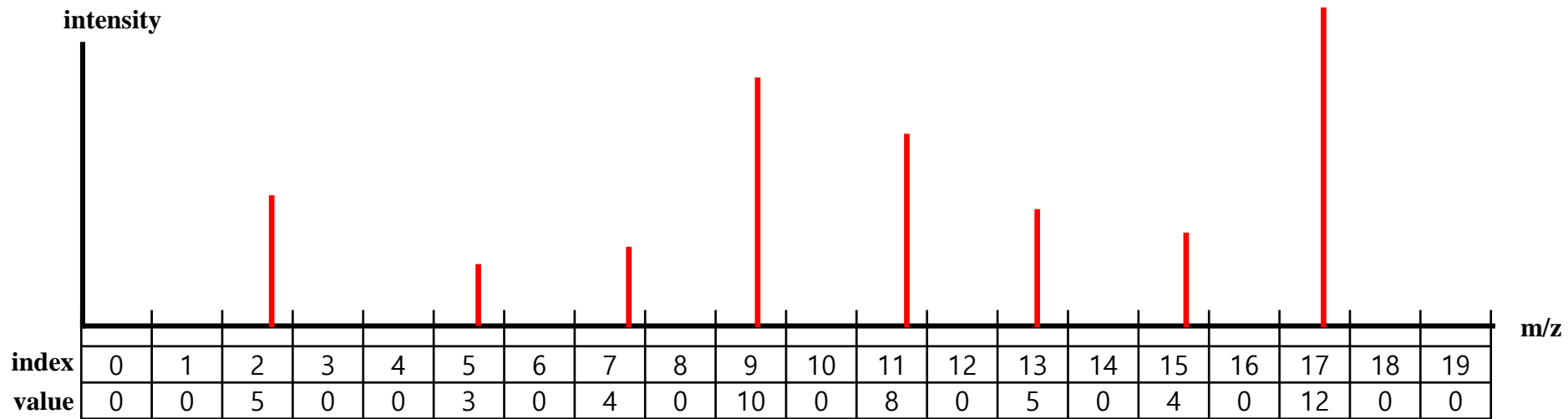
$$\begin{array}{rcl}
 a_0 (b_0 + \cdots + b_{74}) & \longrightarrow & \Theta(\tau) \\
 + a_1 (b_0 + \cdots + b_{74} + b_{75}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{74} (b_0 + \cdots + b_{147} + b_{148}) & \longrightarrow & \Theta(1) \\
 \hline
 + a_{75} (b_1 + \cdots + b_{148} + b_{149}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{n-75} (b_{n-149} + \cdots + b_{n-2} + b_{n-1}) & \longrightarrow & \Theta(1) \\
 \hline
 + a_{n-74} (b_{n-148} + b_{n-147} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1) \\
 \vdots & & \\
 + a_{n-2} (b_{n-76} + b_{n-75} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1) \\
 + a_{n-1} (b_{n-75} + \cdots + b_{n-1}) & \longrightarrow & \Theta(1)
 \end{array}
 \begin{array}{l}
 \left. \begin{array}{l} \text{add} \\ \text{add+subtract} \\ \text{subtract} \end{array} \right\}
 \end{array}$$

Previous Research

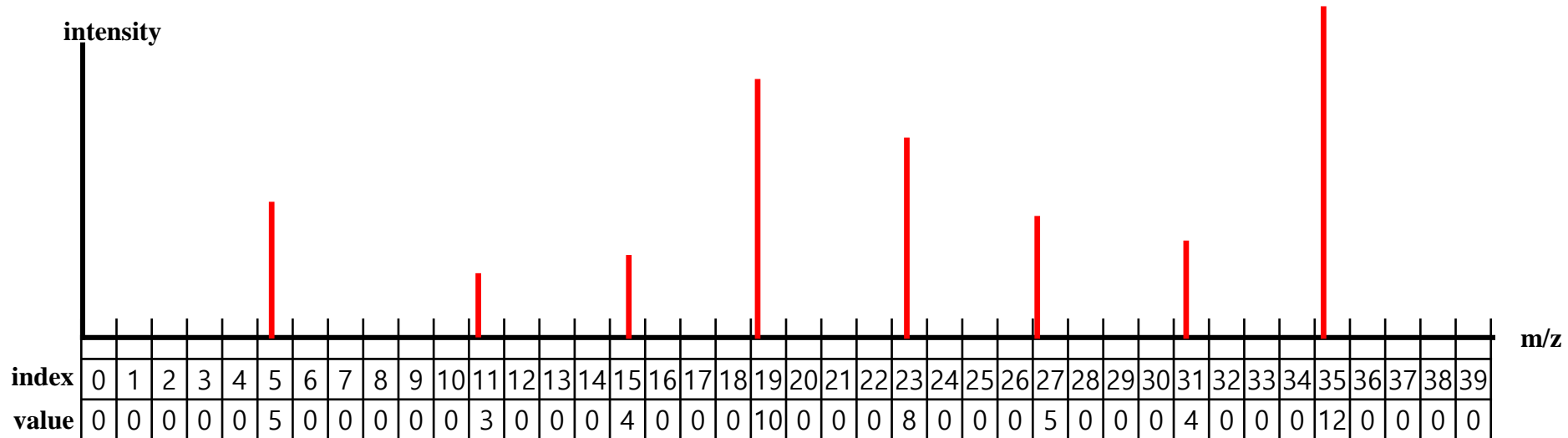
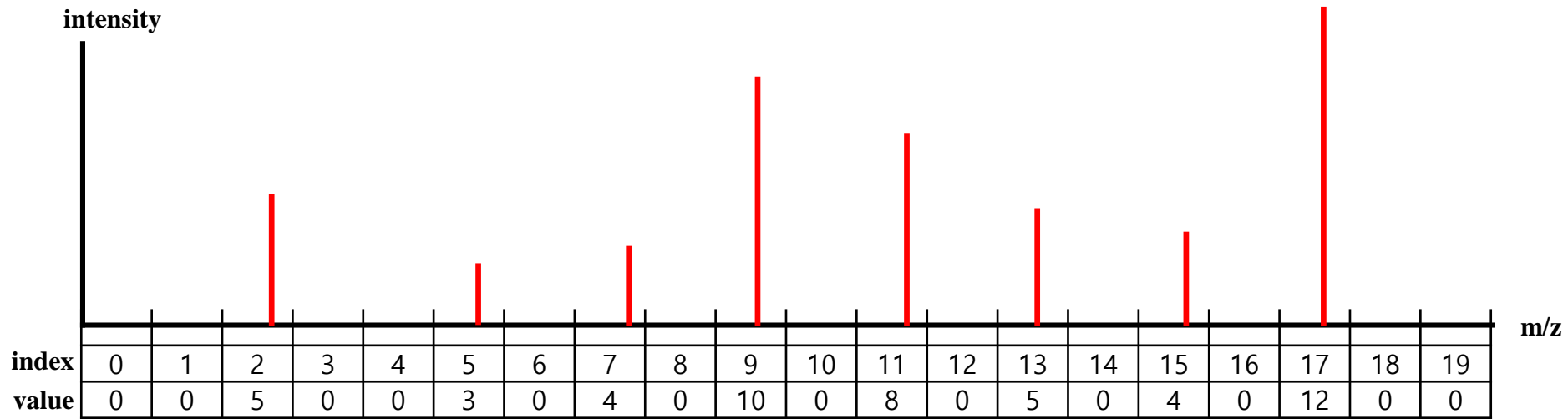
- **Previous algorithm**

- Time complexity is $\Theta(\tau + n)$
 - n is the number of bins in a spectrum
- 최근 장비가 발달함에 따라서 **고해상도 MS/MS가 생산됨**
 - 예전 장비는 저해상도 MS/MS를 생산

Previous Research

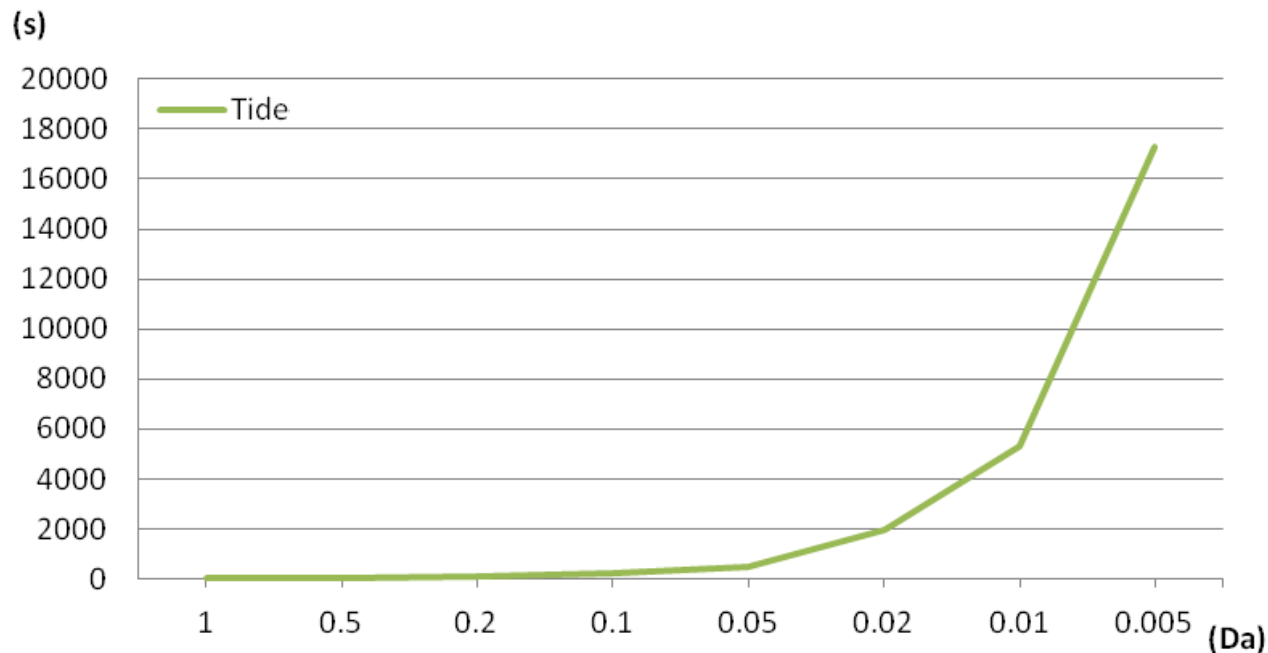


Previous Research



Previous Research

- Measure the running time of Tide

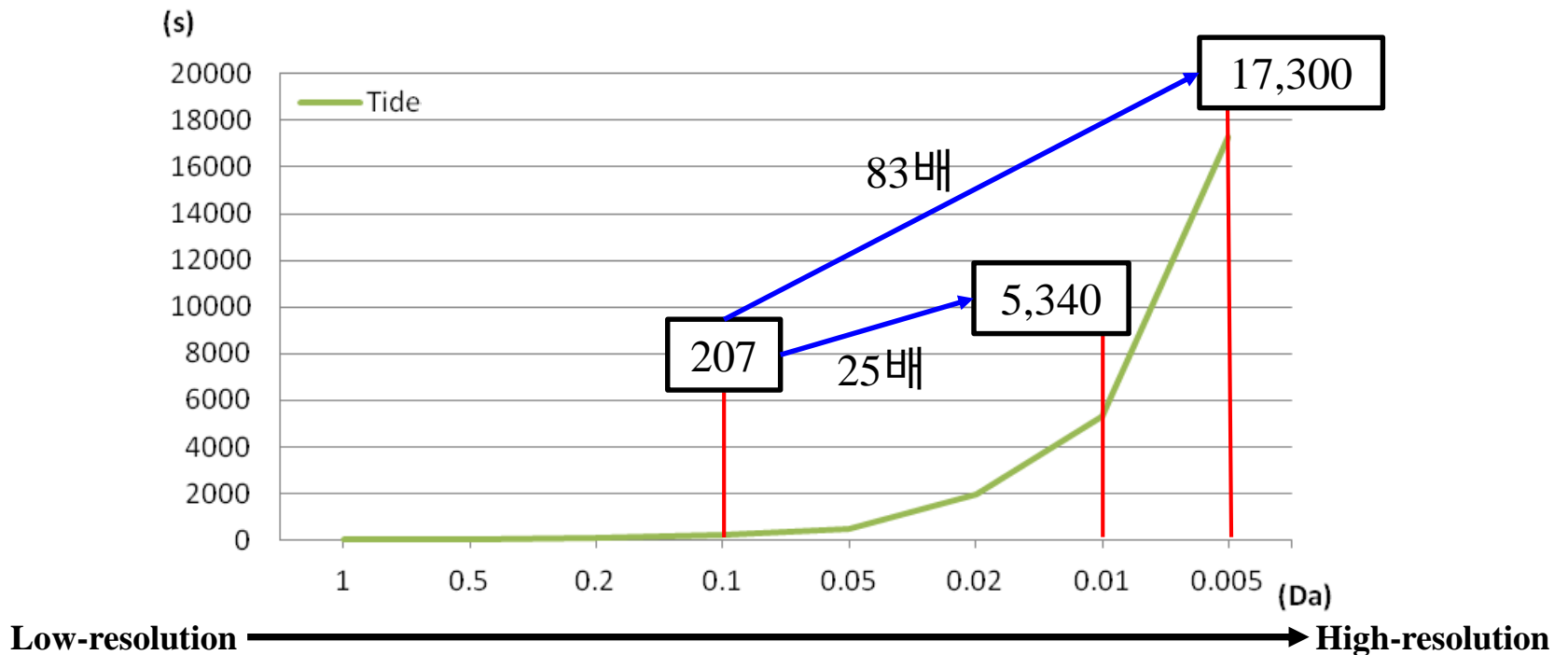


Low-resolution

High-resolution

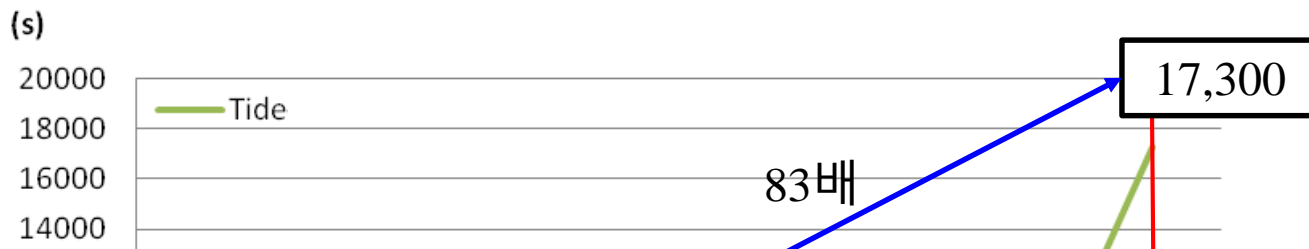
Previous Research

- Measure the running time of Tide



Previous Research

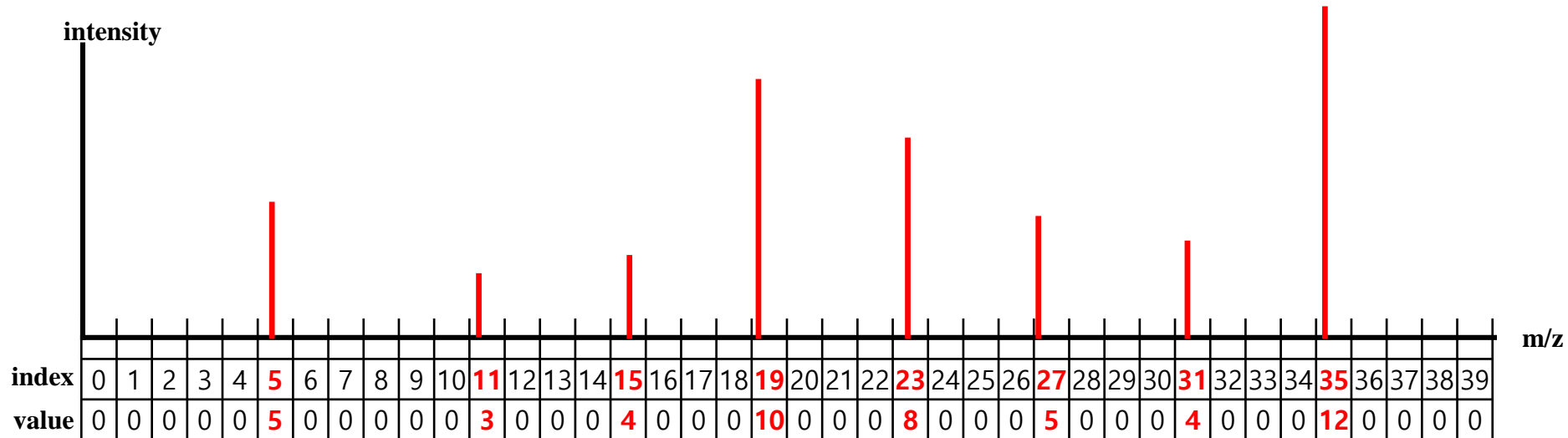
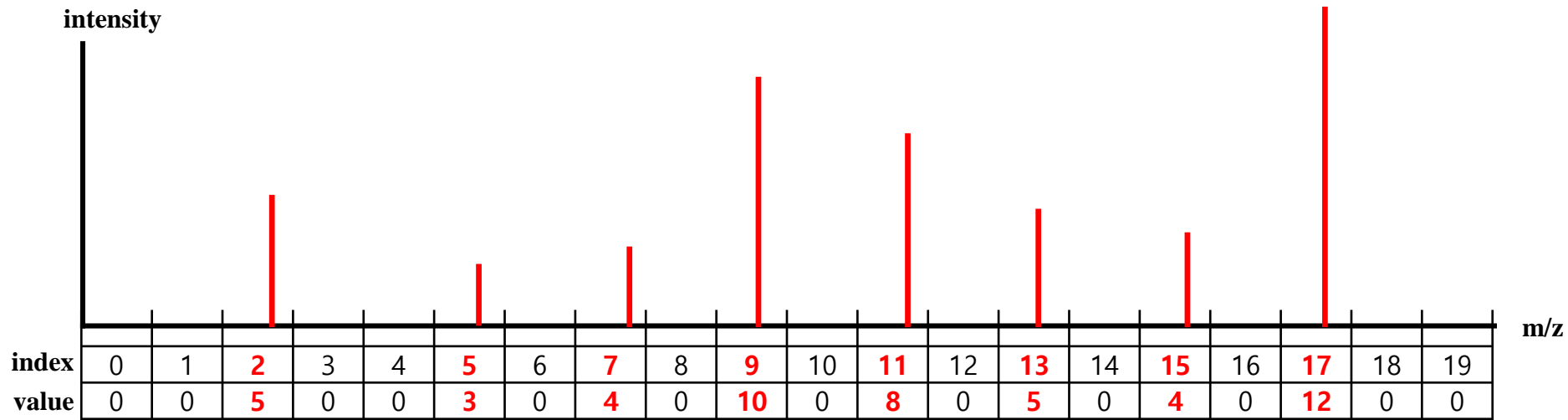
- Measure the running time of Tide



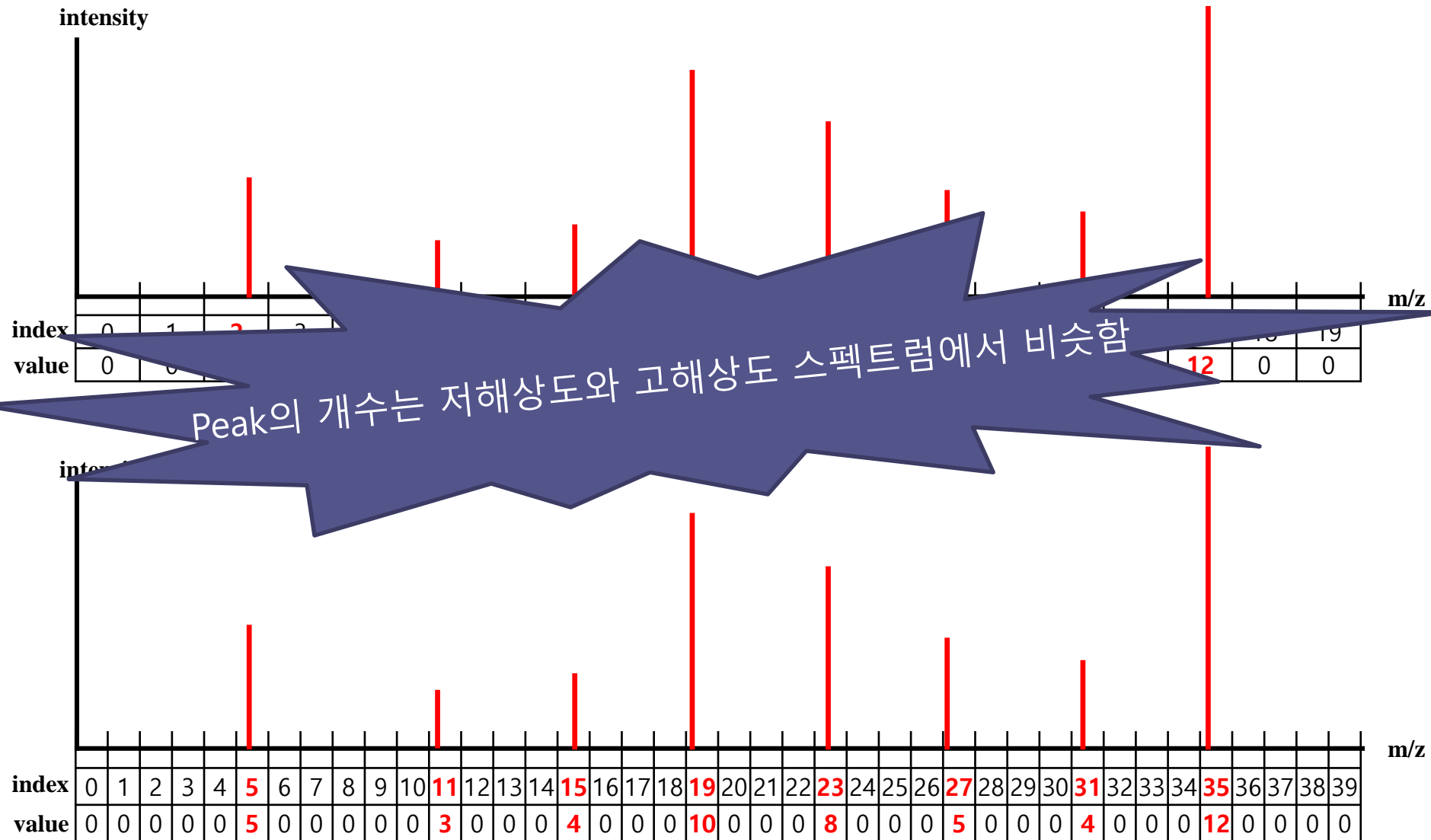
Fragment Tolerance		1	0.5	0.2	0.1	0.05	0.02	0.01	0.005
Tide	Xcorr time	13.51 (50%)	24.43 (58%)	68.77 (69%)	159.1 (76%)	396.8 (82%)	1,727 (87%)	5,313 (99%)	17,208 (99%)
	Total time	26.5	41.9	98.7	207	483	1,970	5,340	17,300

Low-resolution → High-resolution

Previous Research



Previous Research



Our Algorithm

- **Our Algorithm (HiXCorr)**
 - Time complexity is $\Theta(p)$
 - p is the number of peaks in a spectrum

Linear time algorithm

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
value	0	0	5	0	0	3	0	4	0	10	0	8	0	5	0	4	0	12	0	0

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
value	0	0	0	0	0	5	0	0	0	0	0	3	0	0	0	4	0	0	0	10	0	0	0	8	0	0	0	5	0	0	0	4	0	0	0	12	0	0	0	0

0 ($0 + 0 + 5$)
 $+ 0$ ($0 + 0 + 5 + 0$)
 $+ 5$ ($0 + 0 + 5 + 0 + 0$)
 $+ 0$ ($0 + 5 + 0 + 0 + 3$)
 \vdots
 $+ 0$ ($0 + 4 + 0 + 12 + 0$)
 $+ 12$ ($4 + 0 + 12 + 0 + 0$)
 $+ 0$ ($0 + 12 + 0 + 0$)
 $+ 0$ ($12 + 0 + 0$)

0 ($0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0 + 5$)
 $+ 0$ ($0 + 0 + 0 + 5 + 0$)
 $+ 5$ ($0 + 0 + 5 + 0 + 0$)
 \vdots
 $+ 0$ ($0 + 0 + 0 + 12 + 0$)
 $+ 12$ ($0 + 0 + 12 + 0 + 0$)
 $+ 0$ ($0 + 12 + 0 + 0 + 0$)
 $+ 0$ ($12 + 0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0$)

Linear time algorithm

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
value	0	0	5	0	0	3	0	4	0	10	0	8	0	5	0	4	0	12	0	0

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
value	0	0	0	0	0	5	0	0	0	0	0	3	0	0	0	4	0	0	0	10	0	0	0	8	0	0	0	5	0	0	0	4	0	0	0	12	0	0	0	0

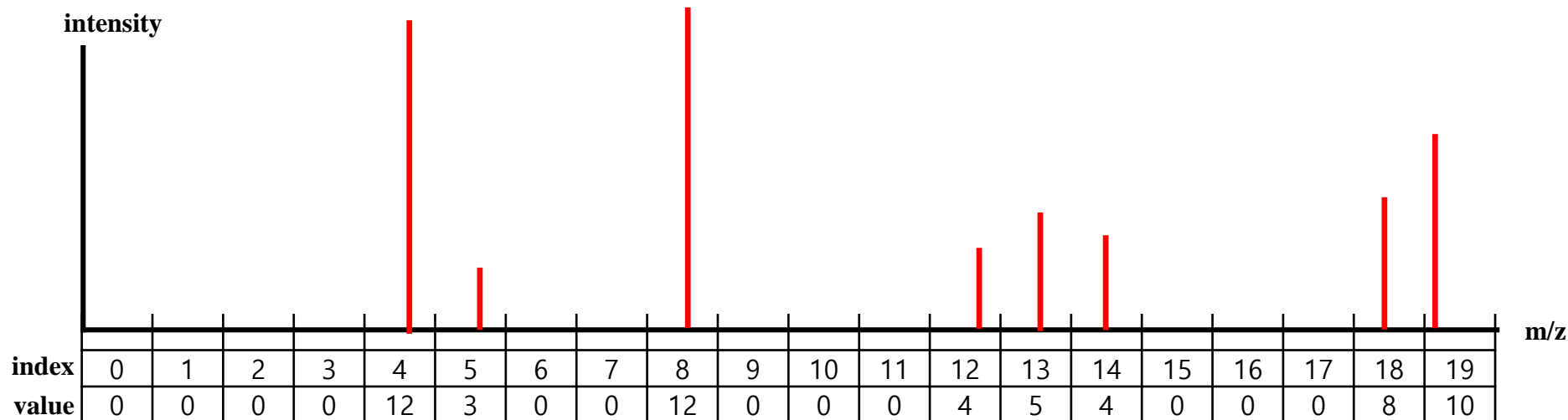
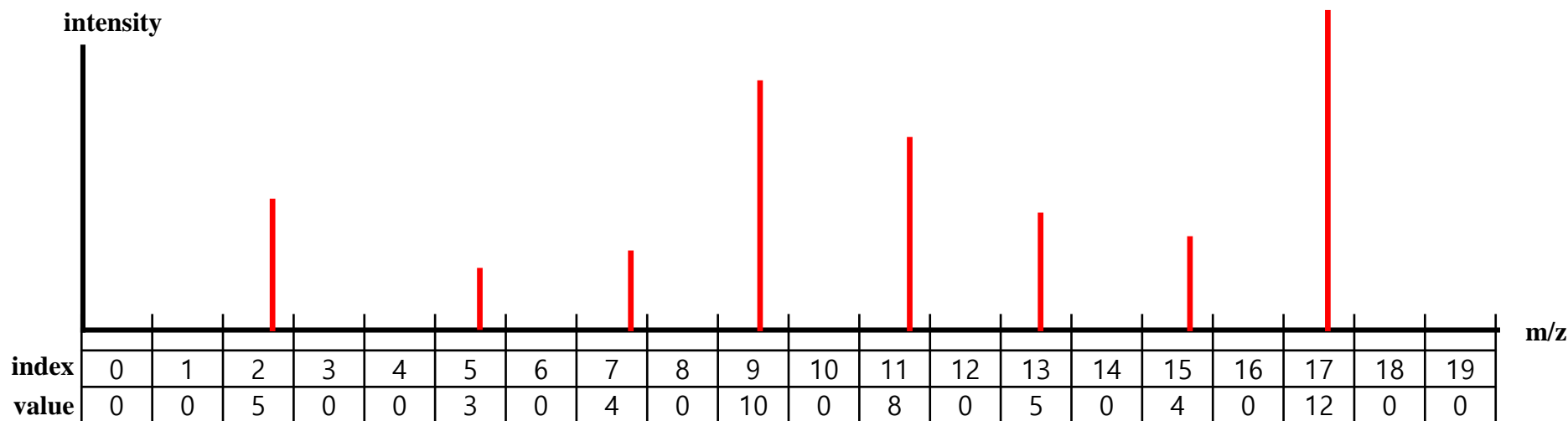
0 ($0 + 0 + 5$)
 $+ 0$ ($0 + 0 + 5 + 0$)
 $+ 5$ ($0 + 0 + 5 + 0 + 0$)
 $+ 0$ ($0 + 5 + 0 + 0 + 3$)
 \vdots
 $+ 0$ ($0 + 4 + 0 + 12 + 0$)
 $+ 12$ ($4 + 0 + 12 + 0 + 0$)
 $+ 0$ ($0 + 12 + 0 + 0$)
 $+ 0$ ($12 + 0 + 0$)

0 ($0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0 + 5$)
 $+ 0$ ($0 + 0 + 0 + 5 + 0$)
 $+ 5$ ($0 + 0 + 5 + 0 + 0$)
 \vdots
 $+ 0$ ($0 + 0 + 0 + 12 + 0$)
 $+ 12$ ($0 + 0 + 12 + 0 + 0$)
 $+ 0$ ($0 + 12 + 0 + 0 + 0$)
 $+ 0$ ($12 + 0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0 + 0$)
 $+ 0$ ($0 + 0 + 0$)

Our Algorithm

- **Our Algorithm (HiXCorr)**
 - Time complexity is $\Theta(p)$
 - p is the number of peaks in a spectrum

Our Algorithm (Example)



Our Algorithm (Example, $\tau=3$)

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0	0	5	0	0	3	0	4	0	10	0	8	0	5	0	4	0	12	0	0
B	0	0	0	0	12	3	0	0	12	0	0	0	4	5	4	0	0	0	8	10

$0 (0 + 0 + 0)$
 $+ 0 (0 + 0 + 0 + 0)$
 $+ 5 (0 + 0 + 0 + 0 + 12)$
 $+ 0 (0 + 0 + 0 + 12 + 3)$
 $+ 0 (0 + 0 + 12 + 3 + 0)$
 $+ 3 (0 + 12 + 3 + 0 + 0)$

 $+ 12 (0 + 0 + 0 + 8 + 10)$
 $+ 0 (0 + 0 + 8 + 10)$
 $+ 0 (0 + 8 + 10)$

Our Algorithm (Example, $\tau=3$)

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0	0	5	0	0	3	0	4	0	10	0	8	0	5	0	4	0	12	0	0
B	0	0	0	0	12	3	0	0	12	0	0	0	4	5	4	0	0	0	8	10

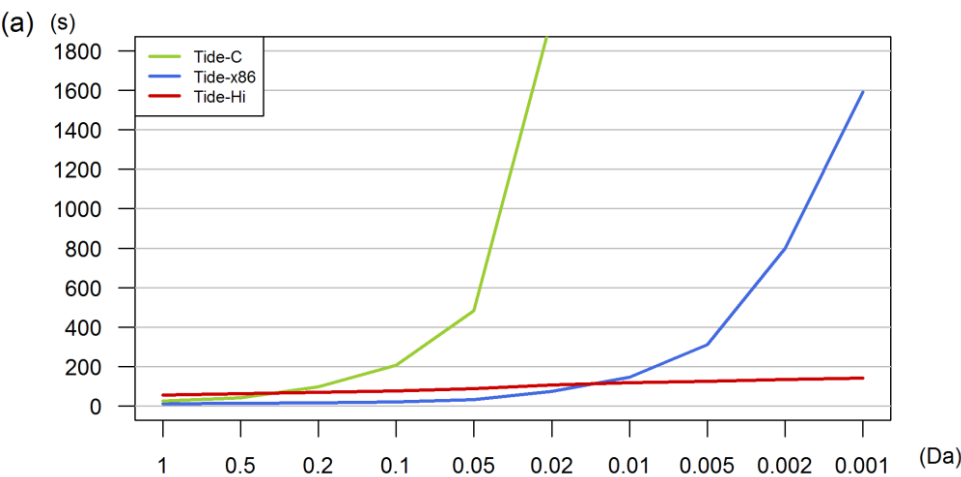
$$\begin{array}{rcl}
 & 0 & (0 + 0 + 0) \\
 + & 0 & (0 + 0 + 0 + 0) \\
 + & 5 & (0 + 0 + 0 + 0 + 12) \\
 + & 0 & (0 + 0 + 0 + 12 + 3) \\
 + & 0 & (0 + 0 + 12 + 3 + 0) \\
 + & 3 & (0 + 12 + 3 + 0 + 0) \\
 \hline
 + & 12 & (0 + 0 + 0 + 8 + 10) \\
 + & 0 & (0 + 0 + 8 + 10) \\
 + & 0 & (0 + 8 + 10)
 \end{array}
 \qquad
 \begin{array}{rcl}
 & + 5 & (12) \\
 & + 3 & (12 + 3) \\
 \hline
 & + 12 & (8 + 10)
 \end{array}$$

Our Algorithm (Example, $\tau=3$)

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	0	0	5	0	0	3	0	4	0	10	0	8	0	5	0	4	0	12	0	0
B	0	0	0	0	12	3	0	0	12	0	0	0	4	5	4	0	0	0	8	10

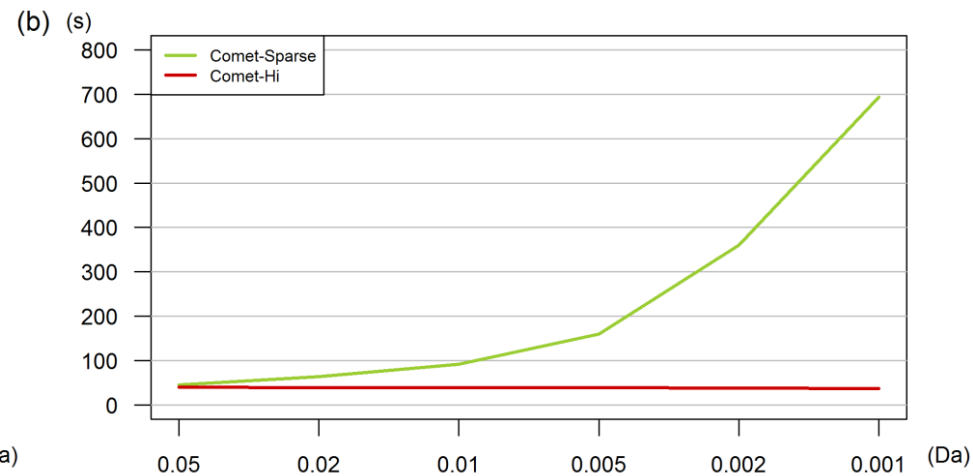
0 (0 + 0 + 0)

+ 0 (0 + 0 + 0 + 0)



+ 0 (0 + 0 + 8 + 10)

+ 0 (0 + 8 + 10)



실습 구현

