# Integrated computational pipeline for the complete sequencing of glycopeptide using HCD/CID-MS/MS and MS3

Mingqi Liu, Wenfeng Zeng, Yang Zhang, Jianqiang Wu, Simin He, Pengyuan Yang

Institutes of Biomedical Sciences, Fudan University, Shanghai
Institute of Computing Technology, Chinese Academy of Sciences, Beijing

CNCP-2014, 11.12-13

Biological function of site-specific glycosylation: Example of IgG
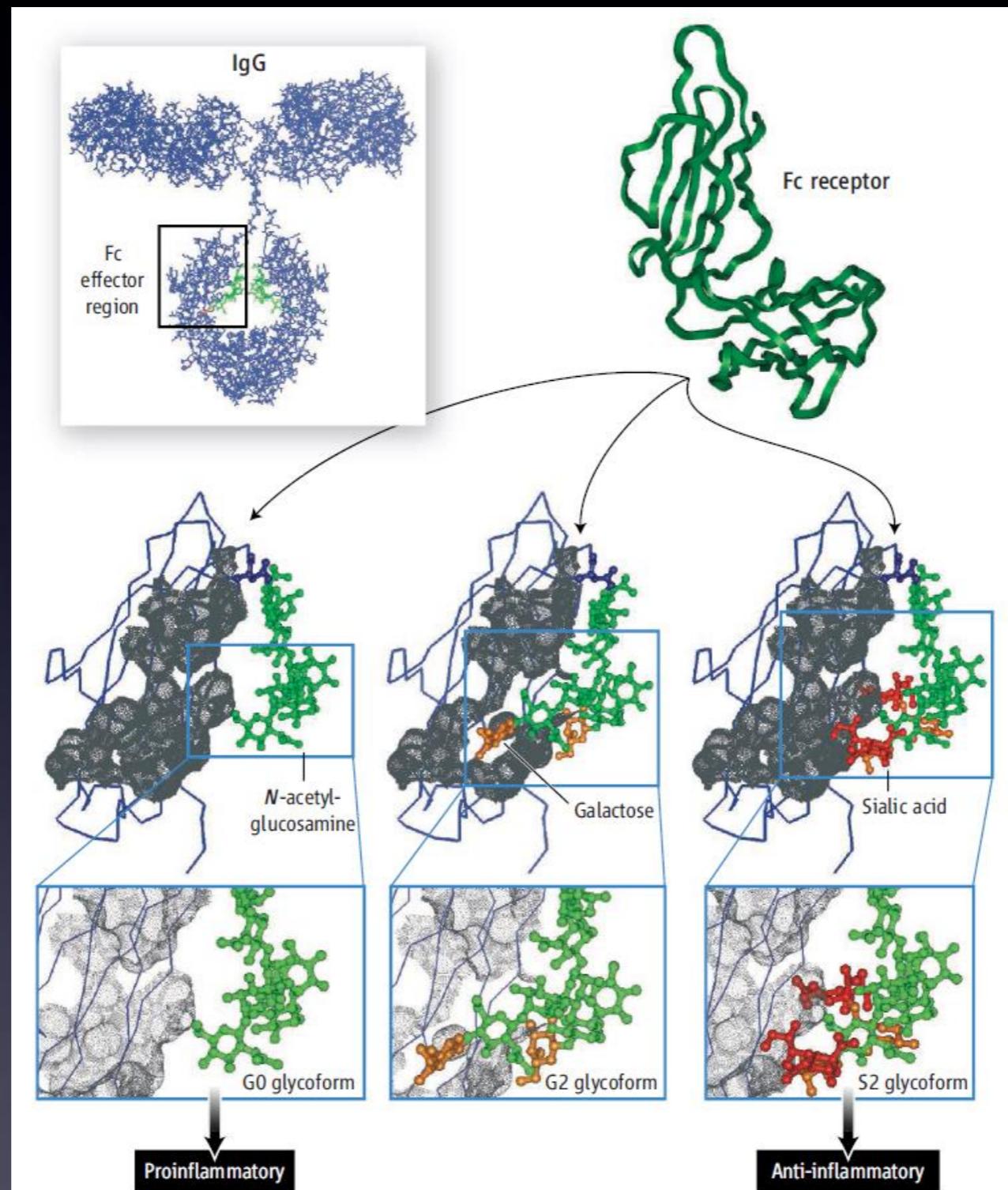
Long term goal:
Interpretation of intact glycopeptide in complex sample

- Study of site-specific glycosylation is developing rapidly

- Identification of intact glycopeptide using LC-MS is the method of choice at the moment

- "Small" obstacle: low abundance, low ionization efficiency

- "Big" obstacle: incomplete information from single MS spectrum

PSM : peptide - spectrum match
"GPSM" : glycopeptide - (2~3) spectra match

Long term goal:
Interpretation of intact glycopeptide in complex sample

- ***Possible strategies***

- Our pipeline

- Challenges

# Complete sequencing of glycopeptide: CID + ETD?

- Typical fragmentation methods:

  CID: collision based, in trap
  HCD: collision based, in beam-type module
 ETD: electric based
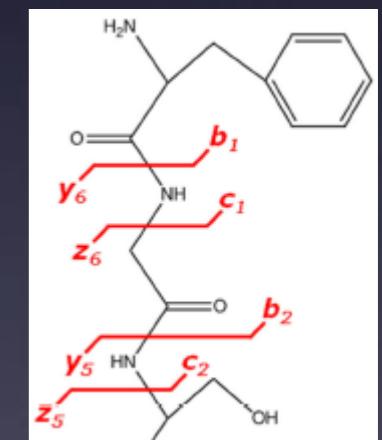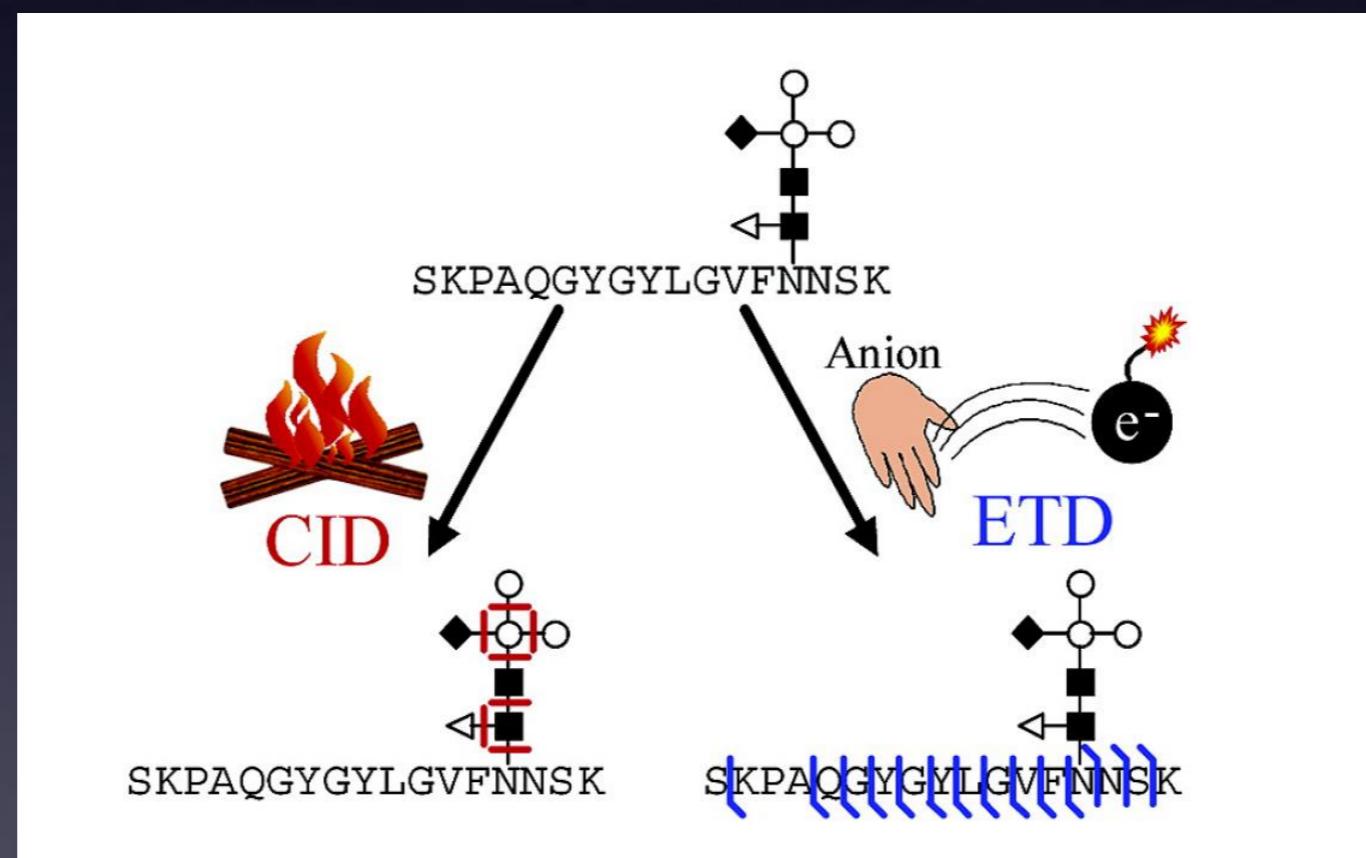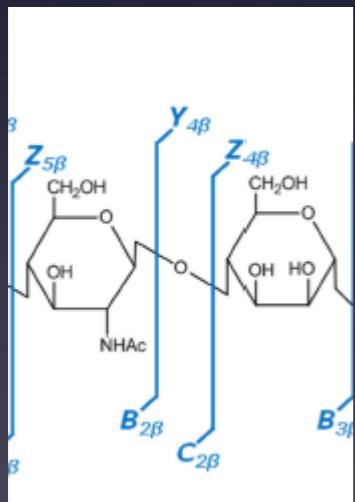
Glycopeptide spectra : ETD
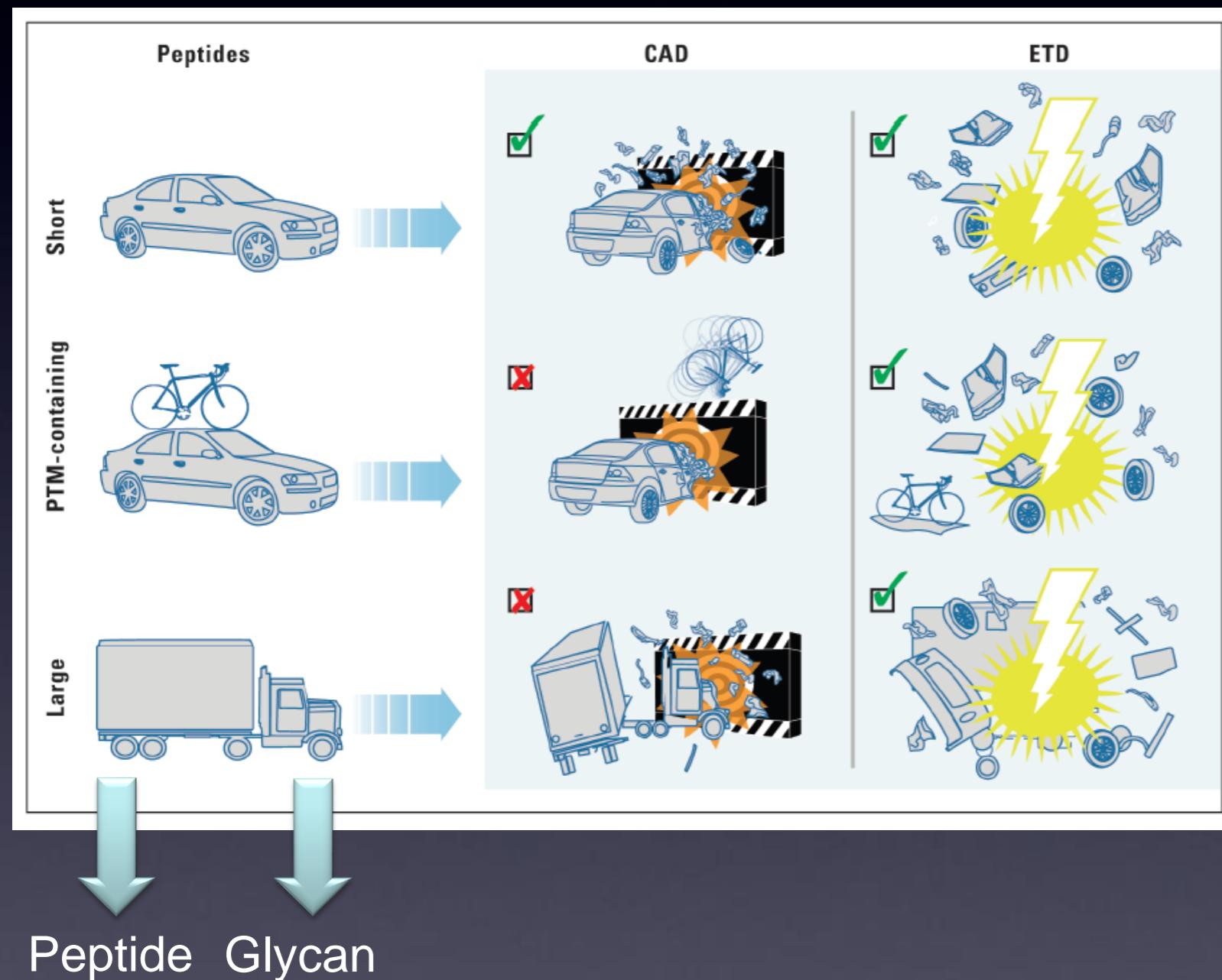
- ETD is widely used in O-glycopeptide profiling. Large-scale N-glycopeptide profiling using ETD is less common.

- It seems that ETD has the "m/z" barrier around 1,000 ~ 1,100. No such barrier is observed in CID.

Glycopeptide spectra : ETD
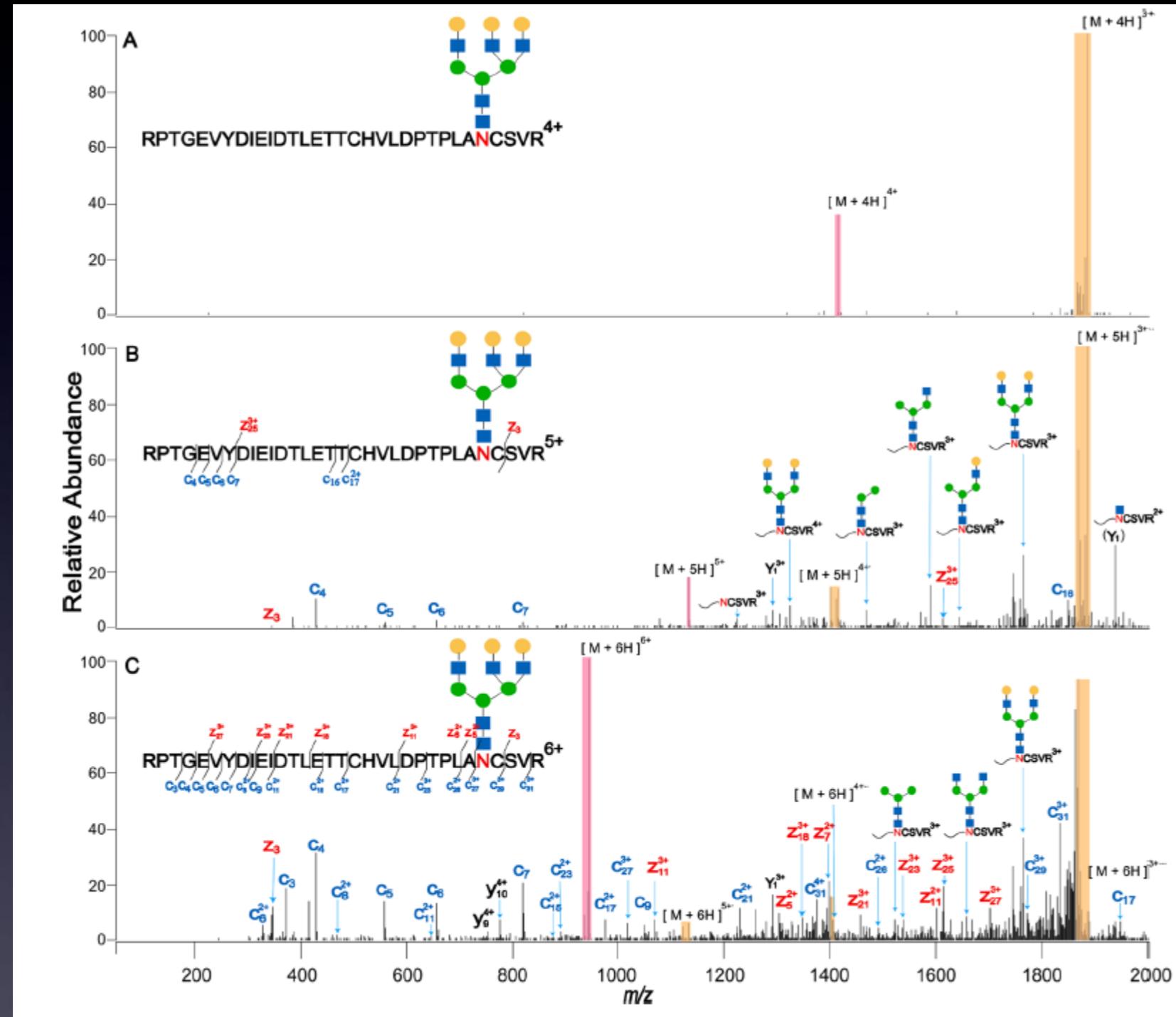
m/z :
1415.6

m/z :
1132.7

m/z :
944.1

Glycopeptide spectra : CID
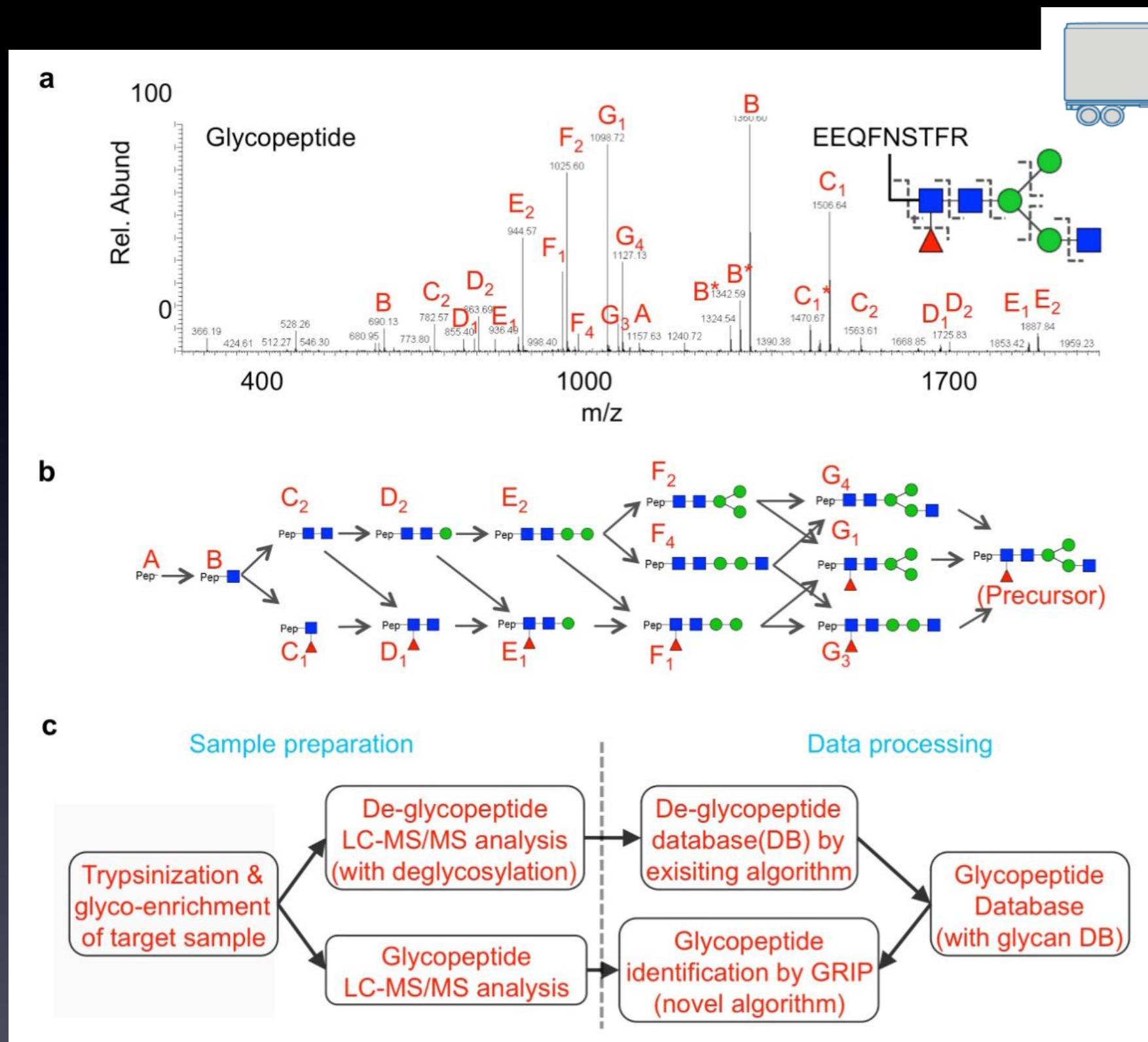
- De-glycosylation is required. Application is limited to N-glycosylation.

- Strictly speaking, information of peptide backbone is incomplete

Long term goal:
Interpretation of intact glycopeptide in complex sample

- Possible strategies

- *Our pipeline*

- Challenges

Integrated computational pipeline for the complete sequencing of glycopeptide using HCD/CID-MS/MS and MS3
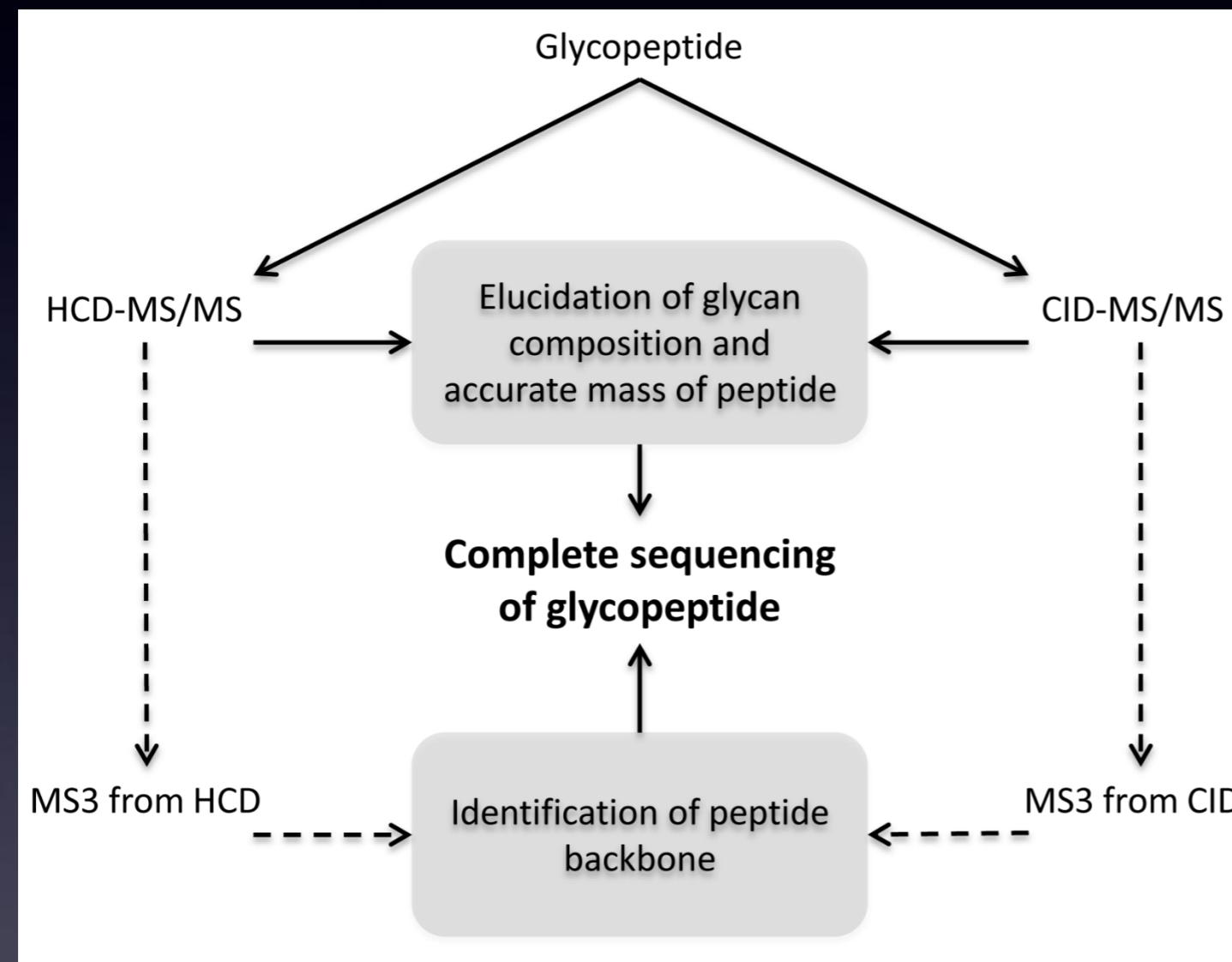
- Sample: Tryptic glycopeptides from mixture of seven standard glycoproteins

- Instrument: Thermo LTQ Orbitrap Fusion

HCD-MS/MS @ 40ev :
Interpretation of the mass of peptide backbone
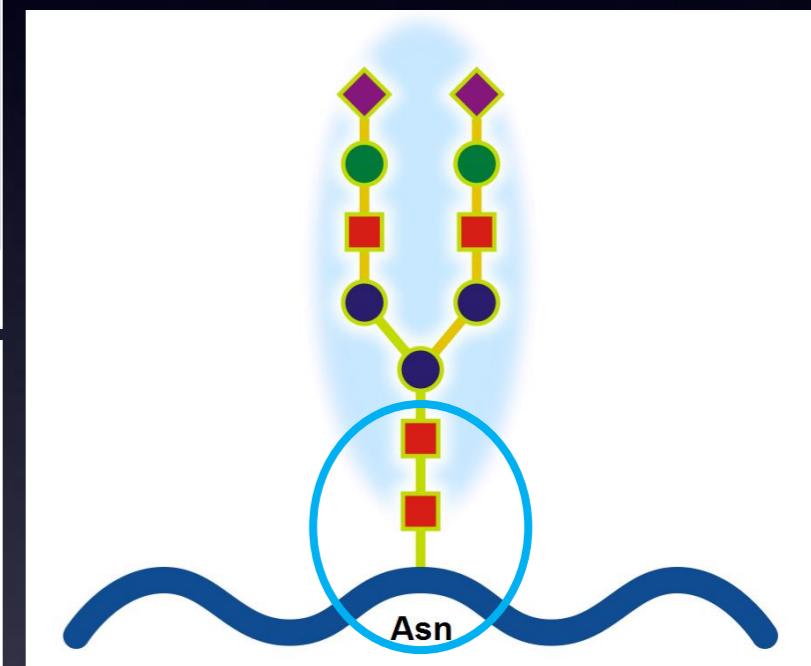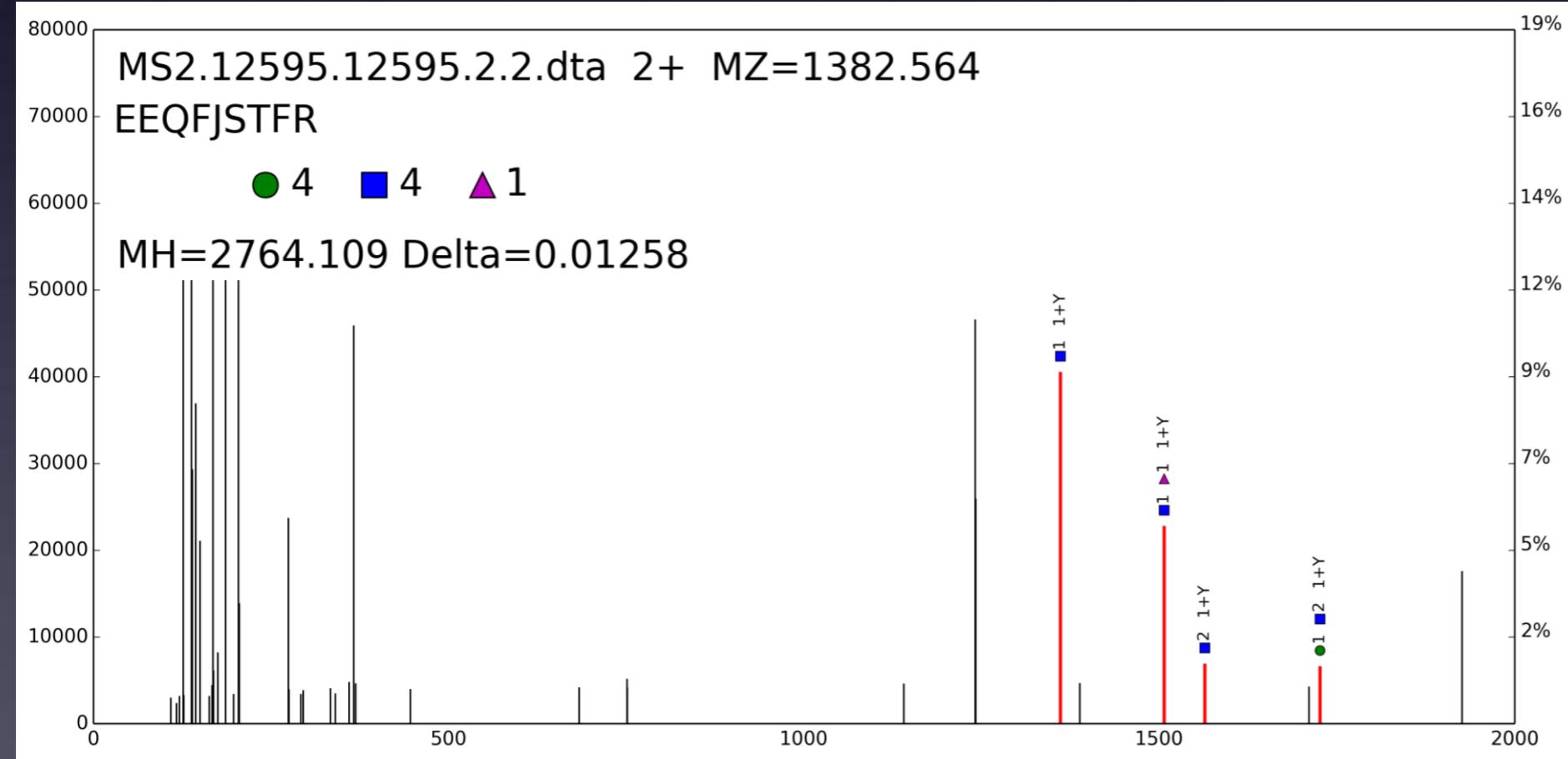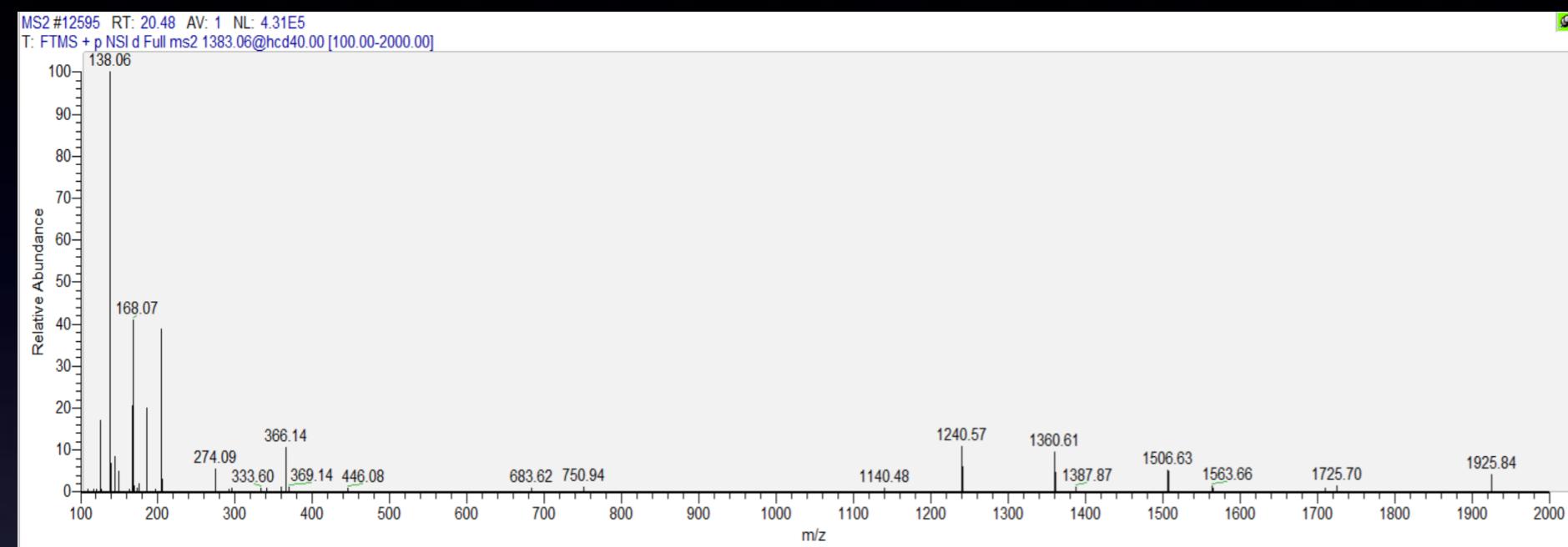
HCD-MS/MS @ 40ev :
Interpretation of the mass of peptide backbone
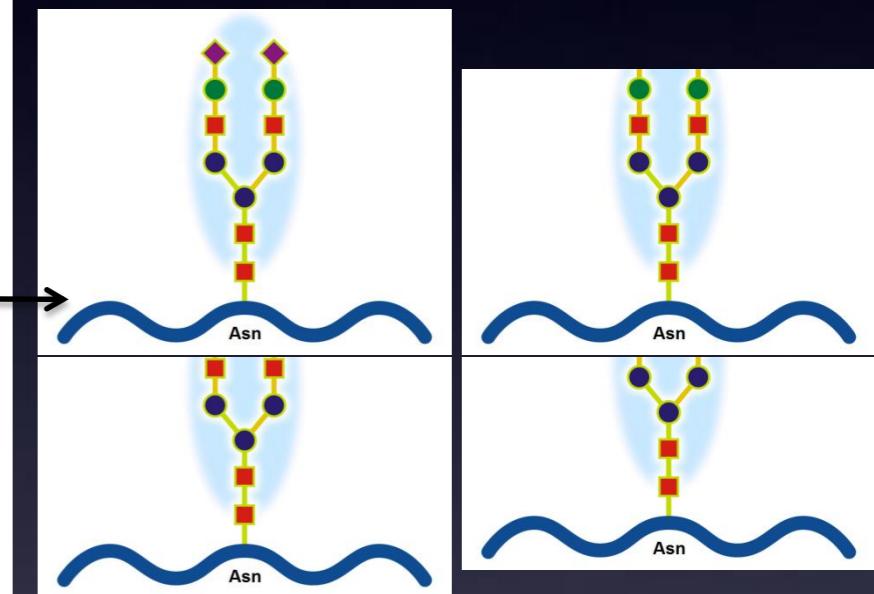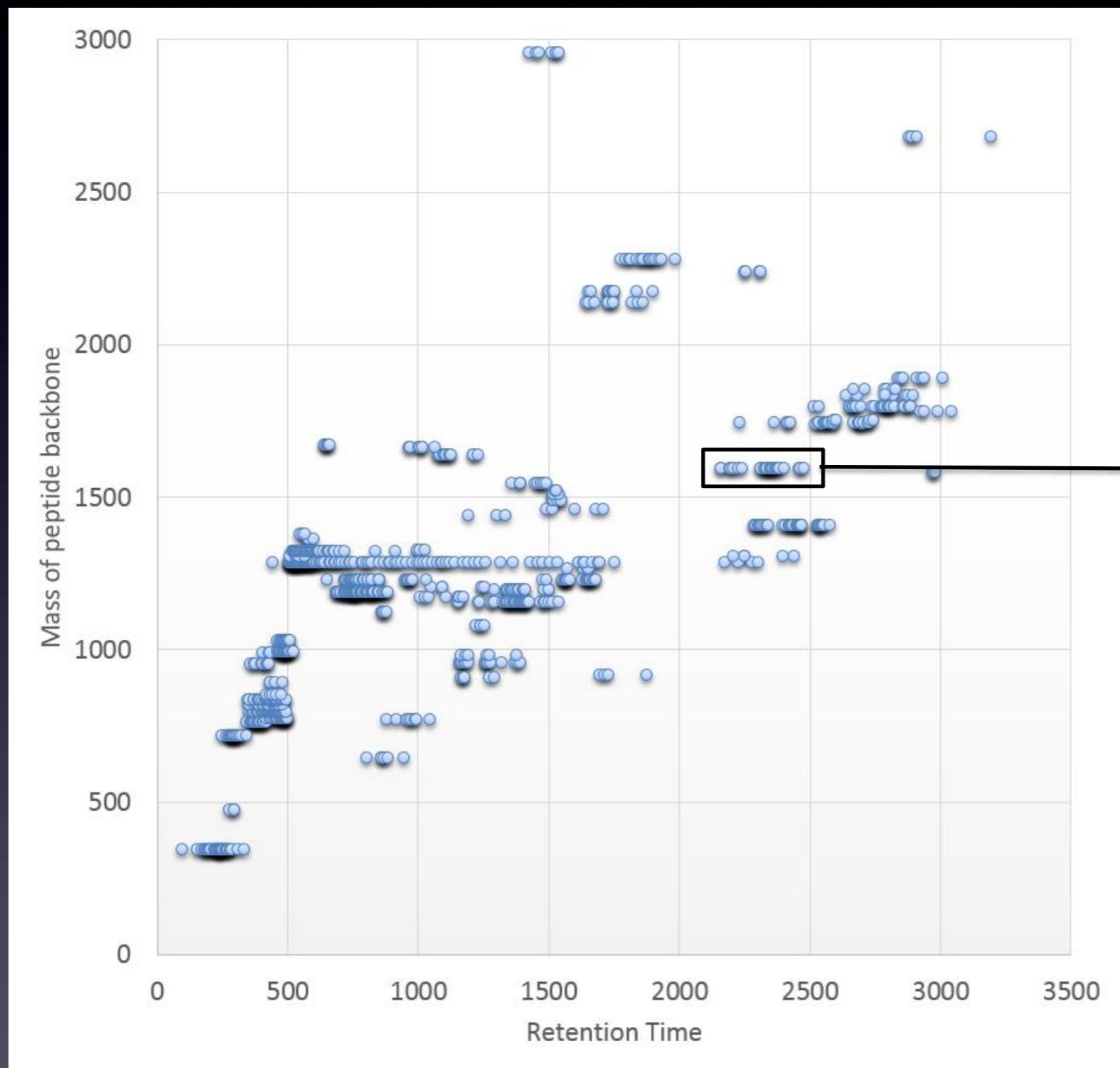
- Interpretation of HCD-MS/MS is protein database free

HCD-MS/MS @ 40ev :
Interpretation of the mass of peptide backbone
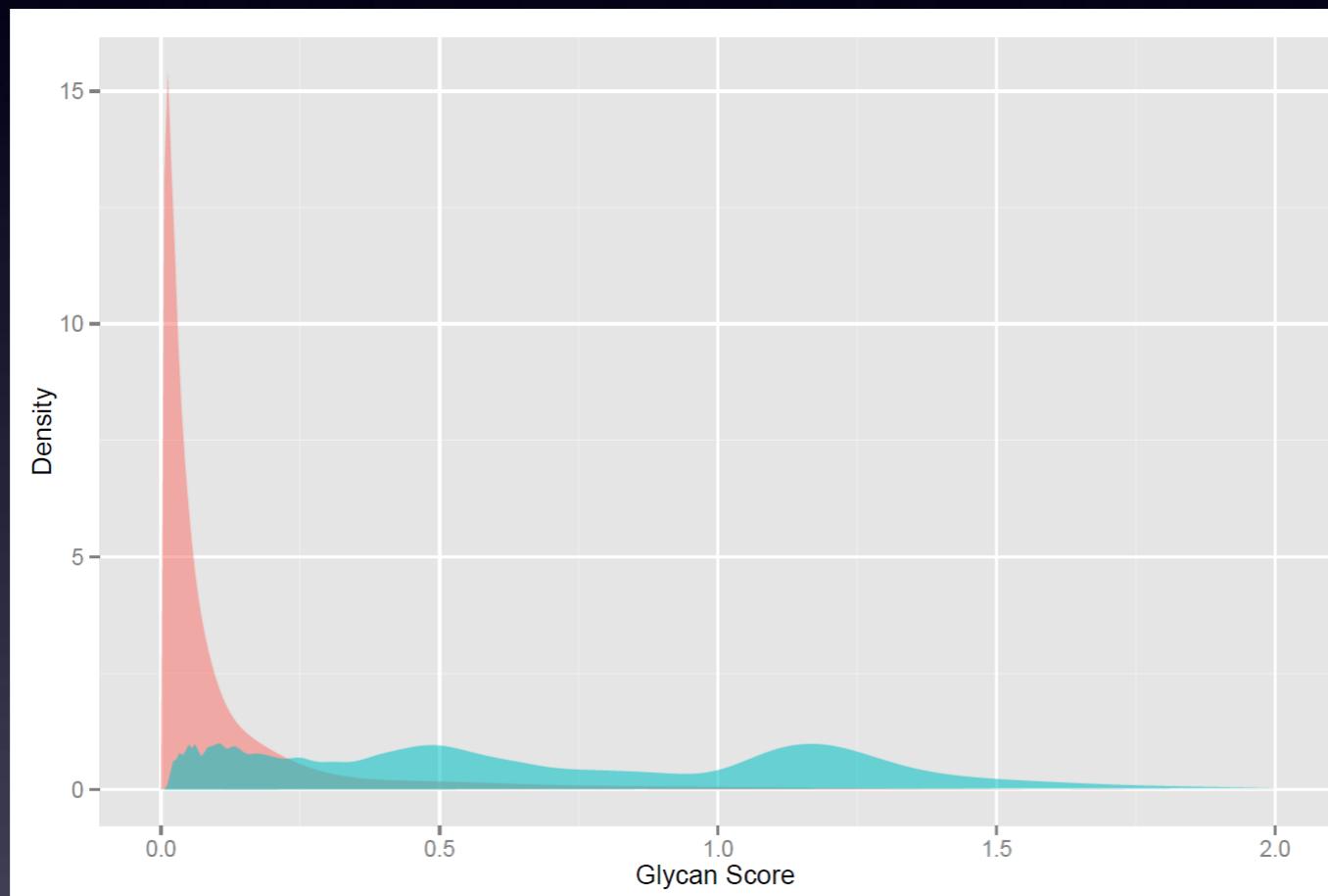
- Target-decoy analysis on HCD-MS/MS. The cutoff we used is 0.4.

CID-MS/MS:
Interpretation of the glycan
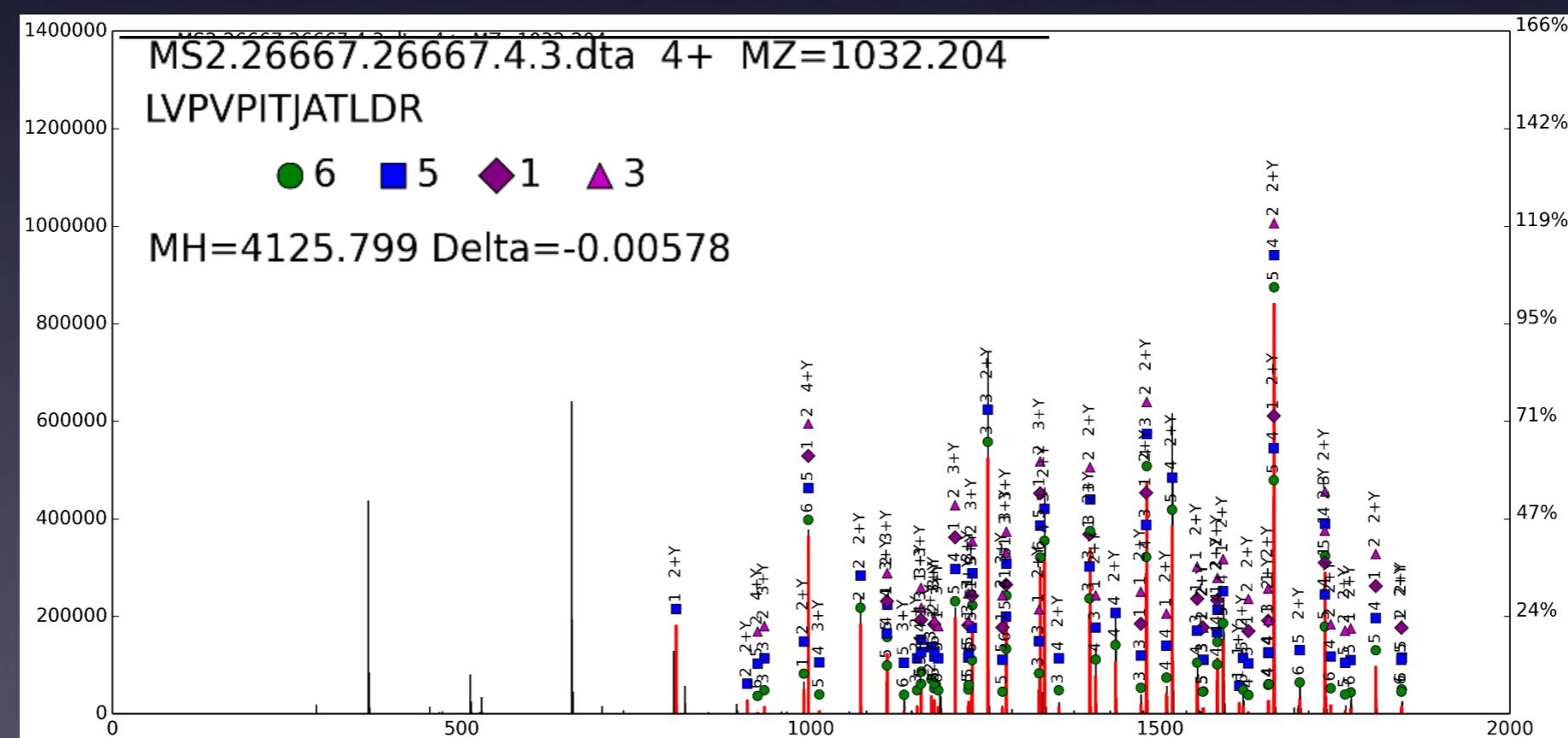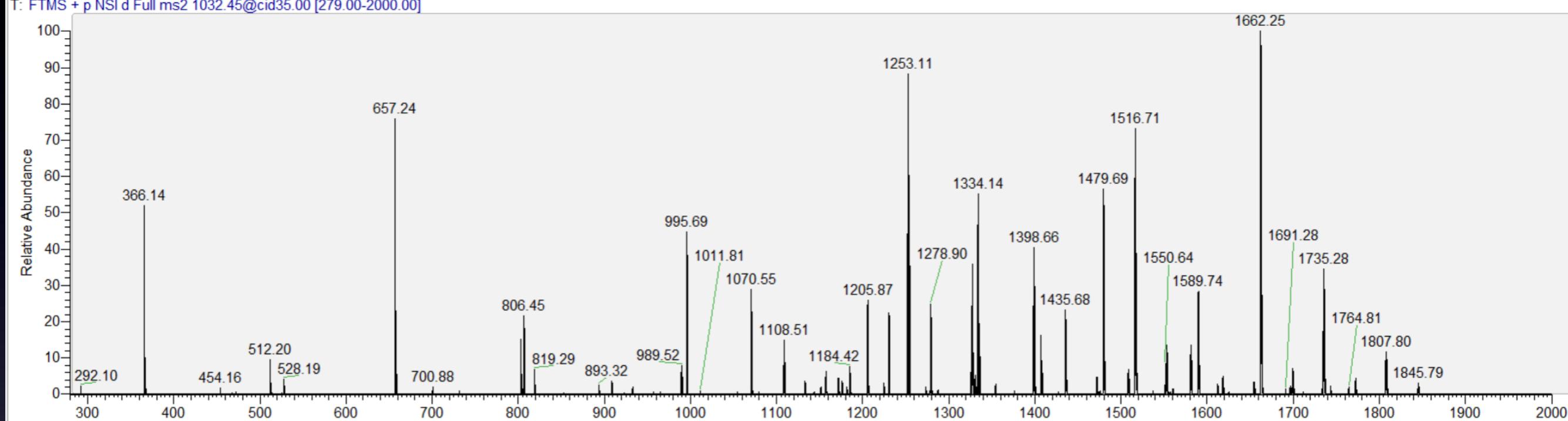
CID-MS/MS:
Interpretation of the glycan
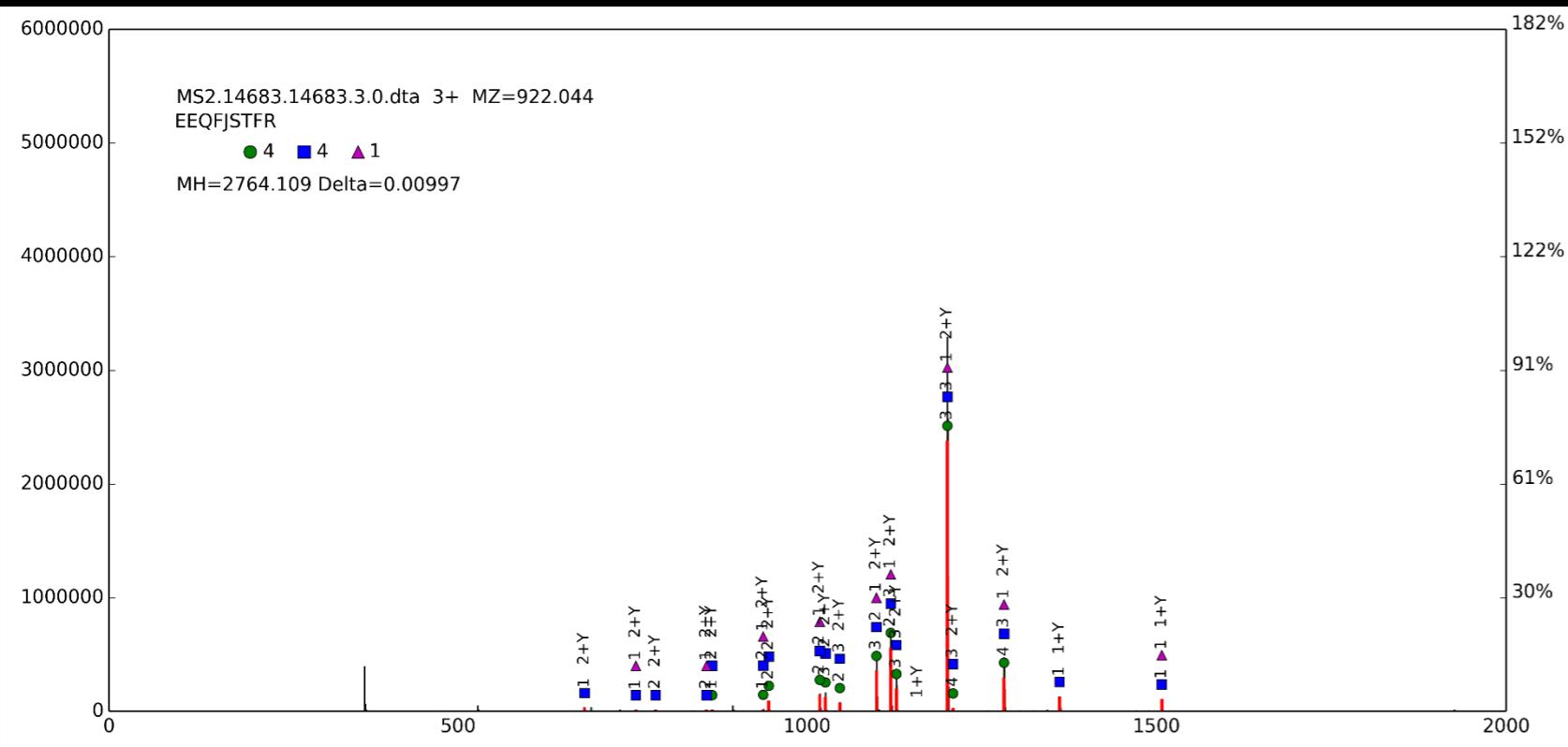
- Glycopeptide CID-MS/MS:

1) Extensive glycosidic bond cleavage

2) Intensity of individual fragments is variable
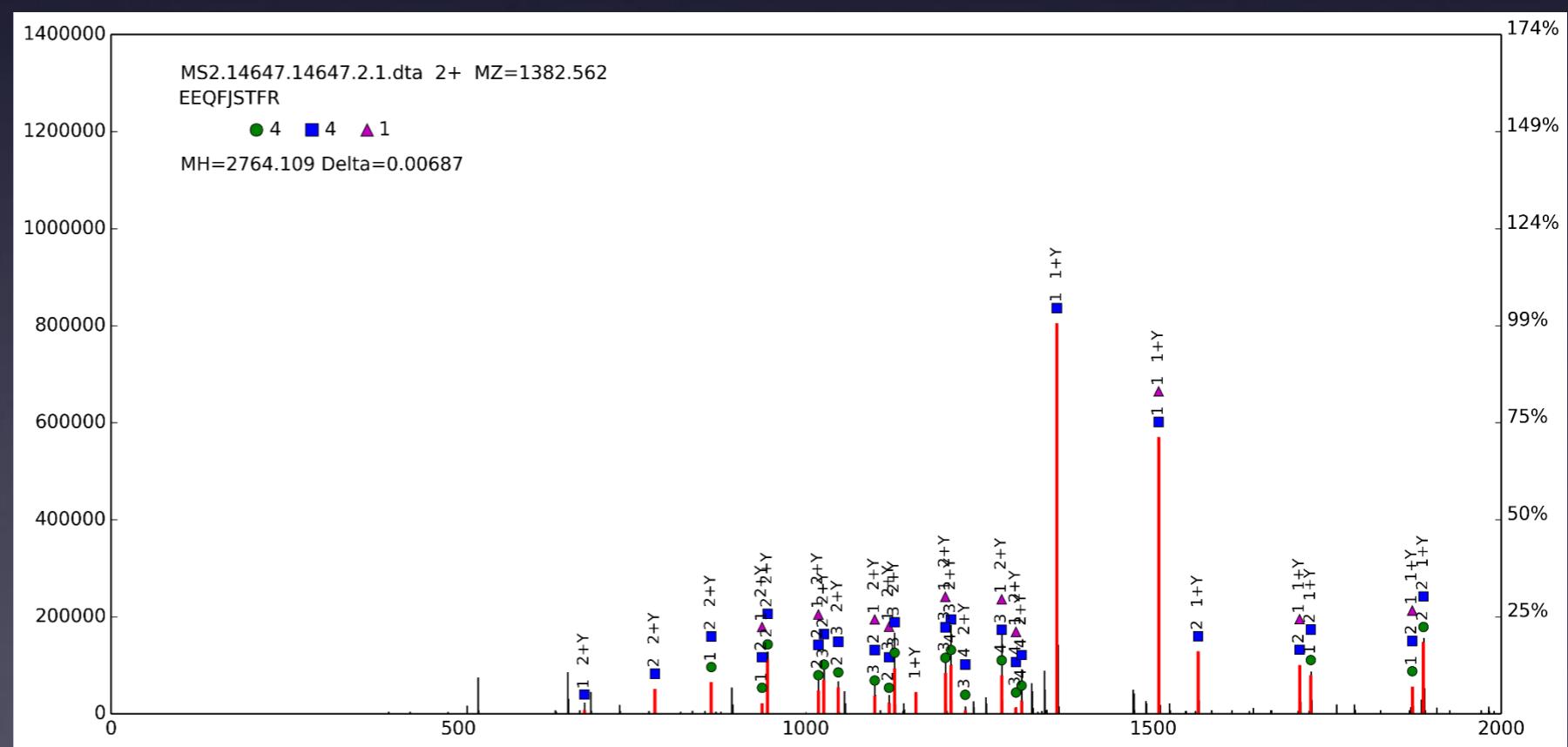
Interpretation of peptide sequence from MS3
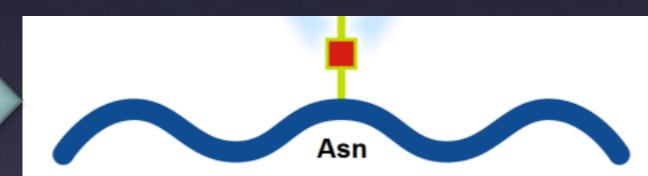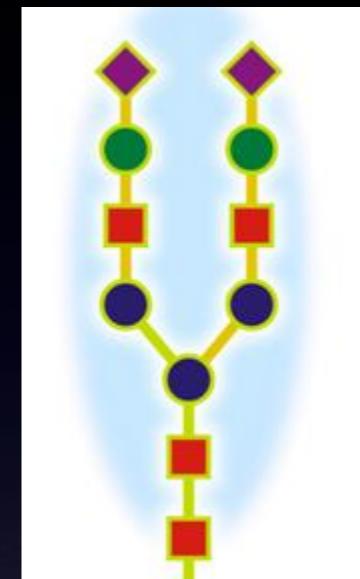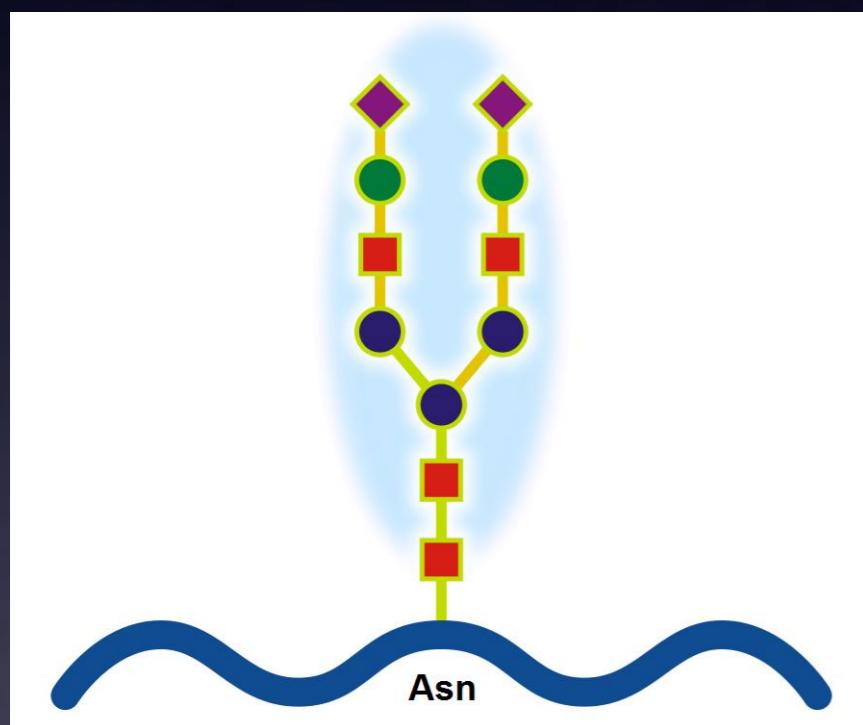
CID-MS2

HCD-MS2@40ev

MS3 from where and how?

Interpretation of peptide sequence from MS3

- Appropriate target and MS parameter for MS3

- We have tested
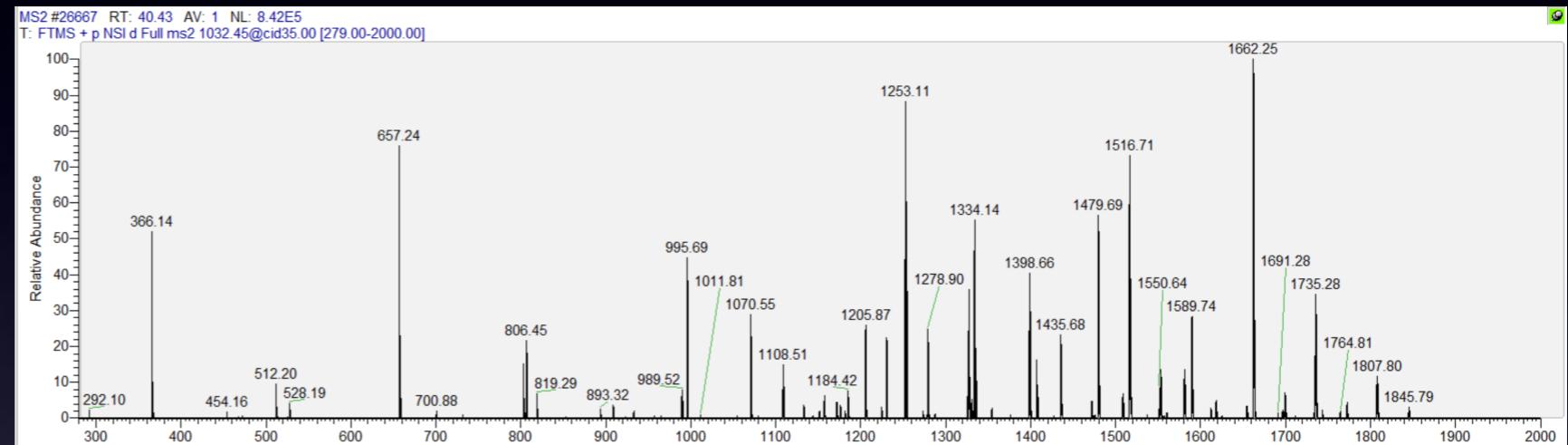
CID-MS/MS-DDA-MS3
CID-MS/MS-DIA-MS3
HCD-MS/MS-DDA-MS3
HCD-MS/MS-DIA-MS3

CID or HCD?
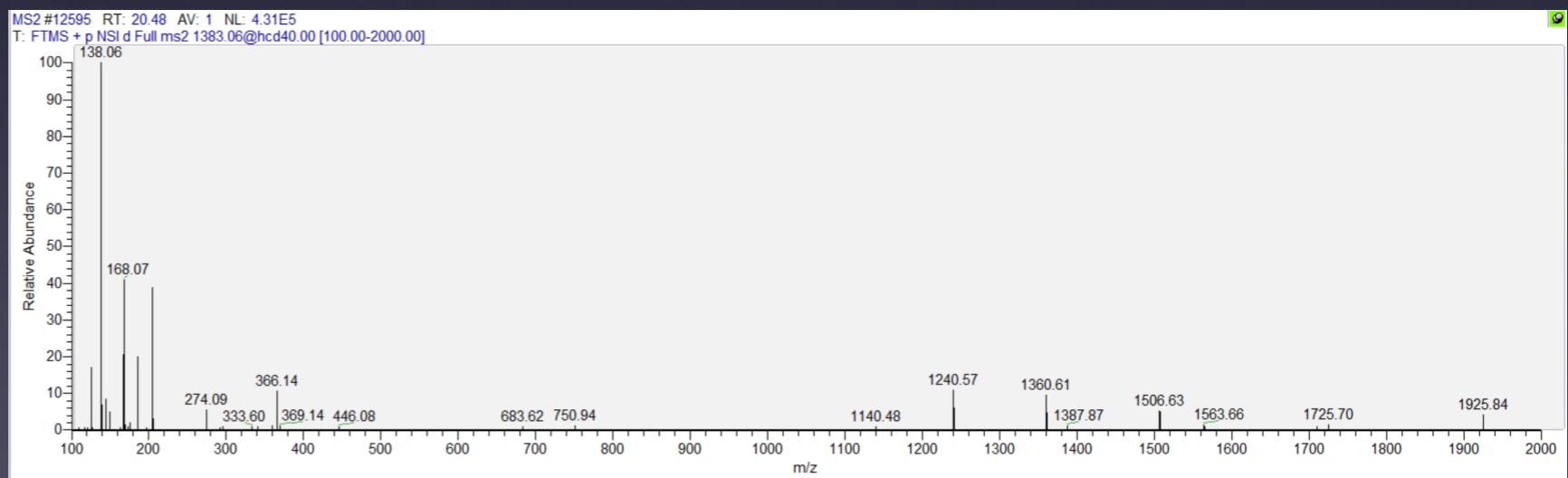
Interpretation of peptide sequence from MS3

- HCD-MS/MS based MS3 can only be performed on Orbitrap Fusion at the moment*

Before Orbitrap Fusion:

OT

Ion source → **IT** → ↑ ↔ → **Q**
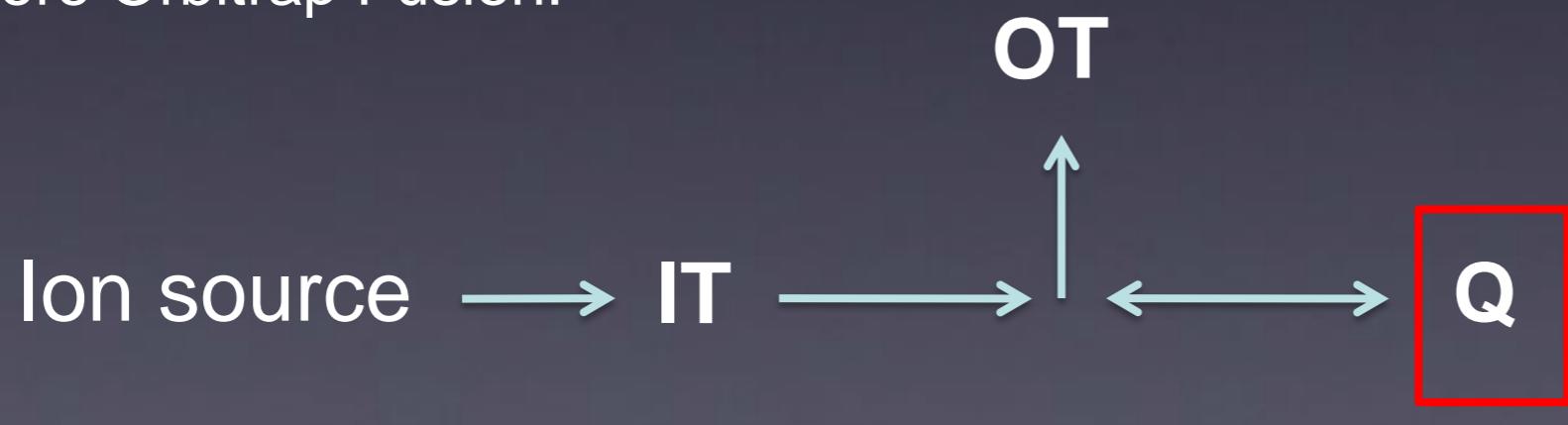
IT: ion trap (CID, MSn)

OT: Orbitrap (Detector)

Q: Quadrupole (HCD)

Orbitrap Fusion:

OT

Ion source → **Q** → ↑ ↔ → **Q** ↔ **IT**
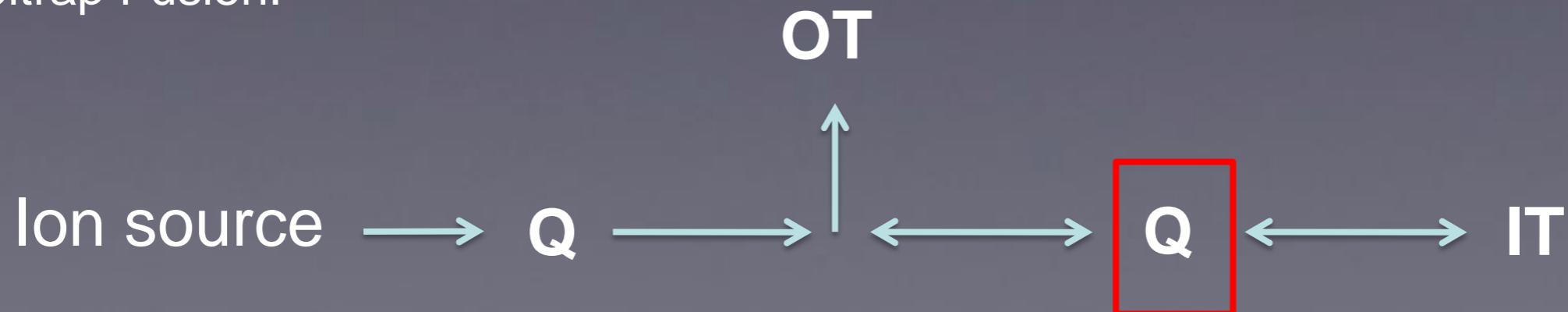
Interpretation of peptide sequence from MS3

- Conventional DDA MS3 from HCD-MS/MS could collect a lot of spectra with peptide sequence information

- Peptide database searching software can be used for peptide sequence in MS3
(pFind, Mascot, SEQUEST, Maxquant, etc.)

- We have develop an integrated software package for the complete sequencing of glycopeptide from MS/MS and MS3

Interpretation of peptide sequence from MS3: Spectral quality

Identified peptide backbone sequence	100 semi	500 semi	2500 semi	All semi
CANLVPVPITJATLDR	Y	Y		
EEQFJSTFR	Y	Y	Y	
EEQFJSTYR	Y	Y		
EEQYJSTYR	Y	Y	Y	
EJGTISR	Y			
EJGTVSR				
FSIJTTNVMGTSLTVR	Y			
FYFTPJK	Y	Y	Y	
FYFTPJKTEDTIFLR	Y	Y	Y	Y
GLTFQQJASSM	Y	Y	Y	Y
GNEANYYSJATTDEHGLVQF	Y	Y		
JNSDISSTR	Y	Y		
JSSYLNVQR	Y			
LAGKPTHVJVS	Y	Y	Y	Y
LAGKPTHVJVSVVM	Y	Y	Y	Y
LAGKPTHVJVSVVMAEVDGTC	Y			
LAGKPTHVJVSVVMAEVDGTCY	Y	Y		
LVPVPITJATLDQITGK	Y	Y	Y	Y
LVPVPITJATLDR	Y	Y	Y	Y
MJVSVCGLYTYGK	Y			
MVSHHJLTTG	Y	Y	Y	
MVSHHJLTTGA	Y	Y	Y	Y
MVSHHJLTTGAT	Y	Y	Y	Y
MVSHHJLTTGATLINEQWLLTTAK	Y	Y		
NEEYJK	Y	Y	Y	Y
NLFLJH	Y	Y	Y	
NLFLJHSEJATAK	Y	Y	Y	Y

- Spectral quality of MS3 was tested by changing the size of protein database*

- The quality MS3 spectra is sufficient for protein database searching

Complete sequencing of glycopeptide

GNR
(RT: 183-334)
54002 / 54003 / 54100 / 54101
54102 / 54103 / 54200 / 54201
55001 / 65103

JNSDISSTR
(RT: 463-501)
43101 / 52000 / 53002 / 53100
53101 / 54000 / 54003 / 54101
55000 / 62000 / 63002 / 63100

EJGTVSR
(RT: 345-410)
54102 / 65102 / 65200 / 75102
76100 / 76200 / 76202 / 76203
76300 / 76301

EEQYJSTYR
(RT: 687-885)
33001 / 34000 / 34001 / 35000
43001 / 44000 / 44001 / 44003
44101 / 45001 / 45101 / 54000
54001 / 54002 / 54003 / 54100
54101 / 54103 / 54201 / 55000
55001 / 55003 / 55101 / 65102

EEQFJSTFR
(RT: 1229-1537)
34001 / 35001 / 44001 / 44003
44100 / 44101 / 45001 / 54001
54003 / 54101

NEEYJK
(RT: 351-496)
54000 / 54002 / 54100 / 54102
54200 / 64102 / 65102 / 65103
65104 / 65200 / 65201 / 65202
65203 / 67104 / 75102 / 75103
75202

YFTPJK
(RT: 918-1047)
54102 / 54200 / 65002 / 65100
65102 / 65200 / 65202 / 75003
75101 / 76200 / 76202 / 76300

EJGTISR
(RT: 434-502)
54102 / 54200 / 65102 / 65104
65200 / 65201 / 65202 / 65300
76100 / 76102 / 76103 / 76104
76200 / 76201 / 76202 / 76300
76301

TPLTAJITK
(RT: 1161-1392)
45001 / 45101 / 54001 / 54002
54100 / 54101 / 54103 / 54201
55001 / 55003 / 55101 / 55201
65102

0   900   1800
Retention Time

LVPVPITJATLDR
(RT: 2287-2576)
54100 / 54102 / 54200 / 64101
65002 / 65003 / 65004 / 65100
65101 / 65102 / 65103 / 65200
65201 / 65300 / 75001 / 75002
75101 / 75102 / 75103 / 75201
76002 / 76200

VVLHPJYSQVDIGLIK
(RT: 2515-2886)
54000 / 54002 / 54100 / 54102
54200 / 54201 / 64002 / 64003
64101 / 65000 / 65004 / 65100
65102 / 65103 / 65104 / 65200
65201 / 65202 / 65300 / 75101
75202 / 76004 / 76102 / 76200

SIJTTNVMGTSLTVR
(RT: 2197-2464)
43002 / 43100 / 44100 / 53101
54000 / 54001 / 54002 / 54100
54101 / 63002 / 64001 / 64002
76005

YFTPJKTEDTIFLR
(RT: 2368-2722)
54100 / 65002 / 65004 / 65100
65103 / 65104 / 65200 / 65201
65202 / 65203 / 65300 / 75101
75102 / 75201 / 75202 / 76003
76004 / 76101 / 76104 / 76200
76202 / 76300 / 76301

1800   2400   3600
**Retention Time (Continued)**

- For each glycopeptide identification, a pair of HCD/CID-MS/MS is needed

- A group of glycopeptide identification shares one or more MS3 spectra for peptide backbone sequencing
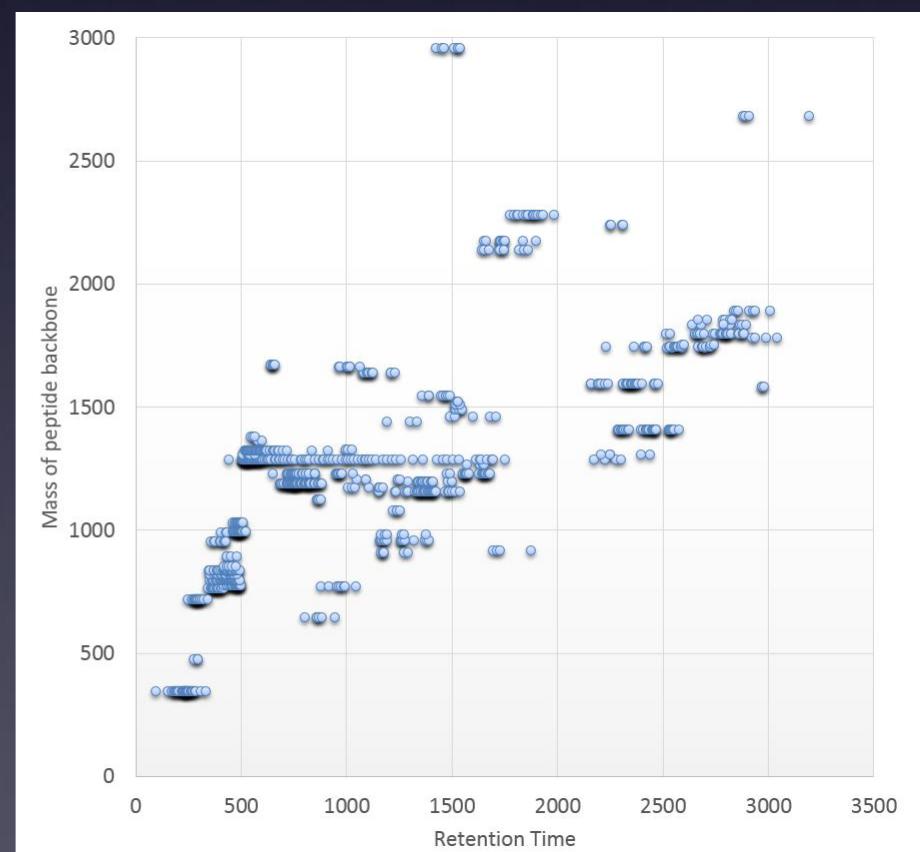
Exemplary spectra:
CID-MS/MS

MS2.4873.4873.3.0.dta 3+ MZ=1115.461

YKJNSDISSTR

●5 ■4 ◆1 ▲1

MH=3344.353 Delta=0.01581
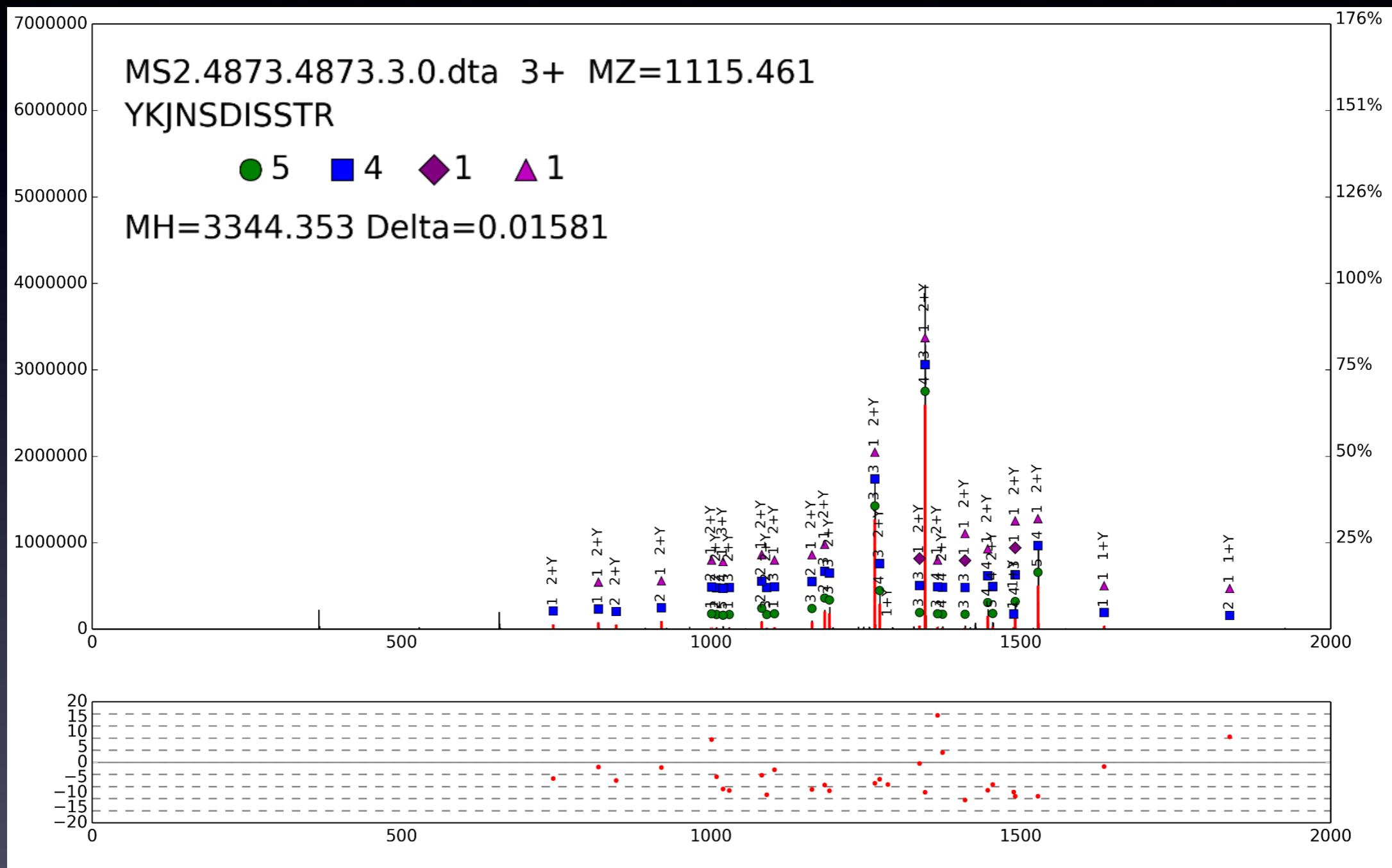
Exemplary spectra:
HCD-MS/MS
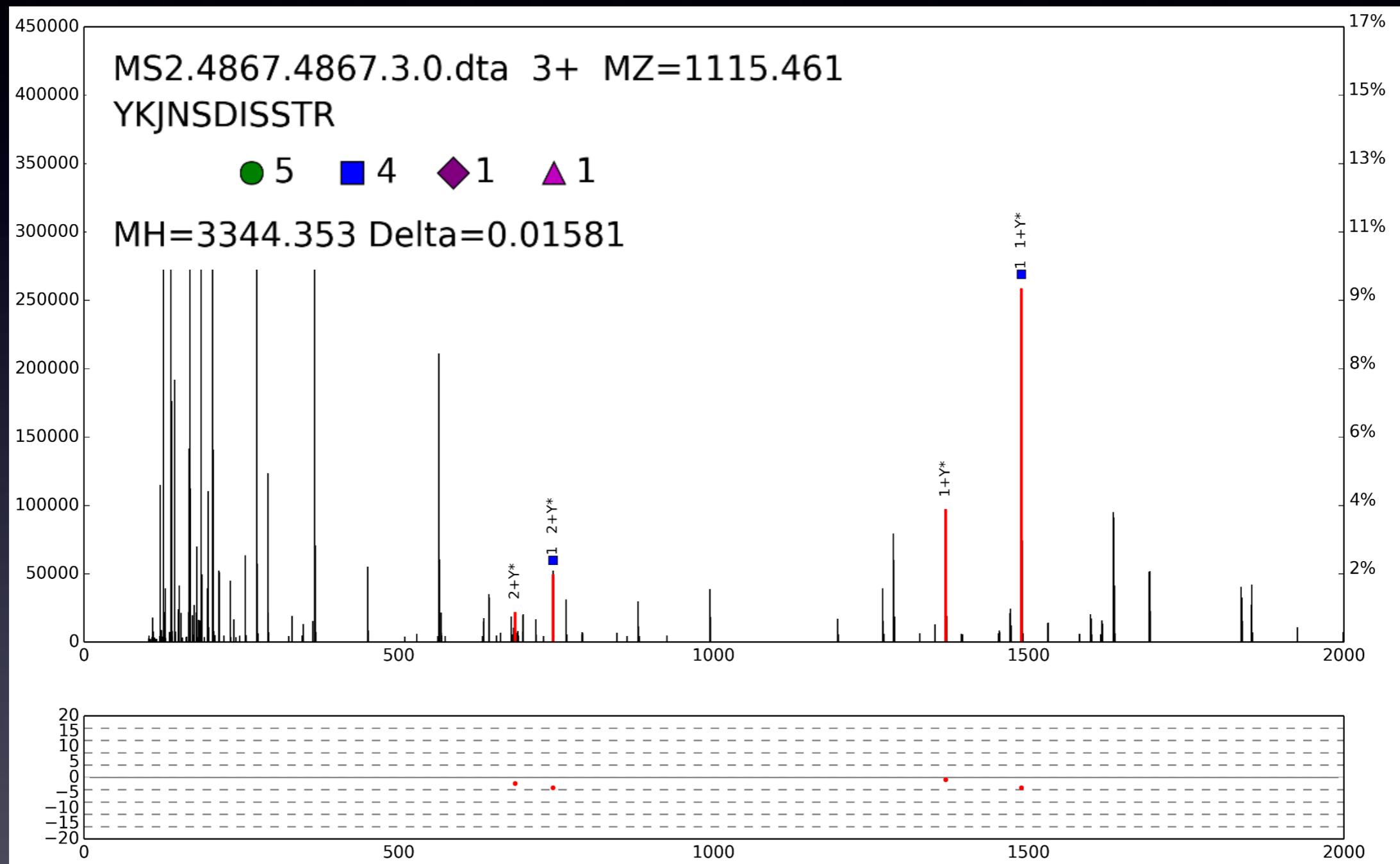
Exemplary spectra:
MS3
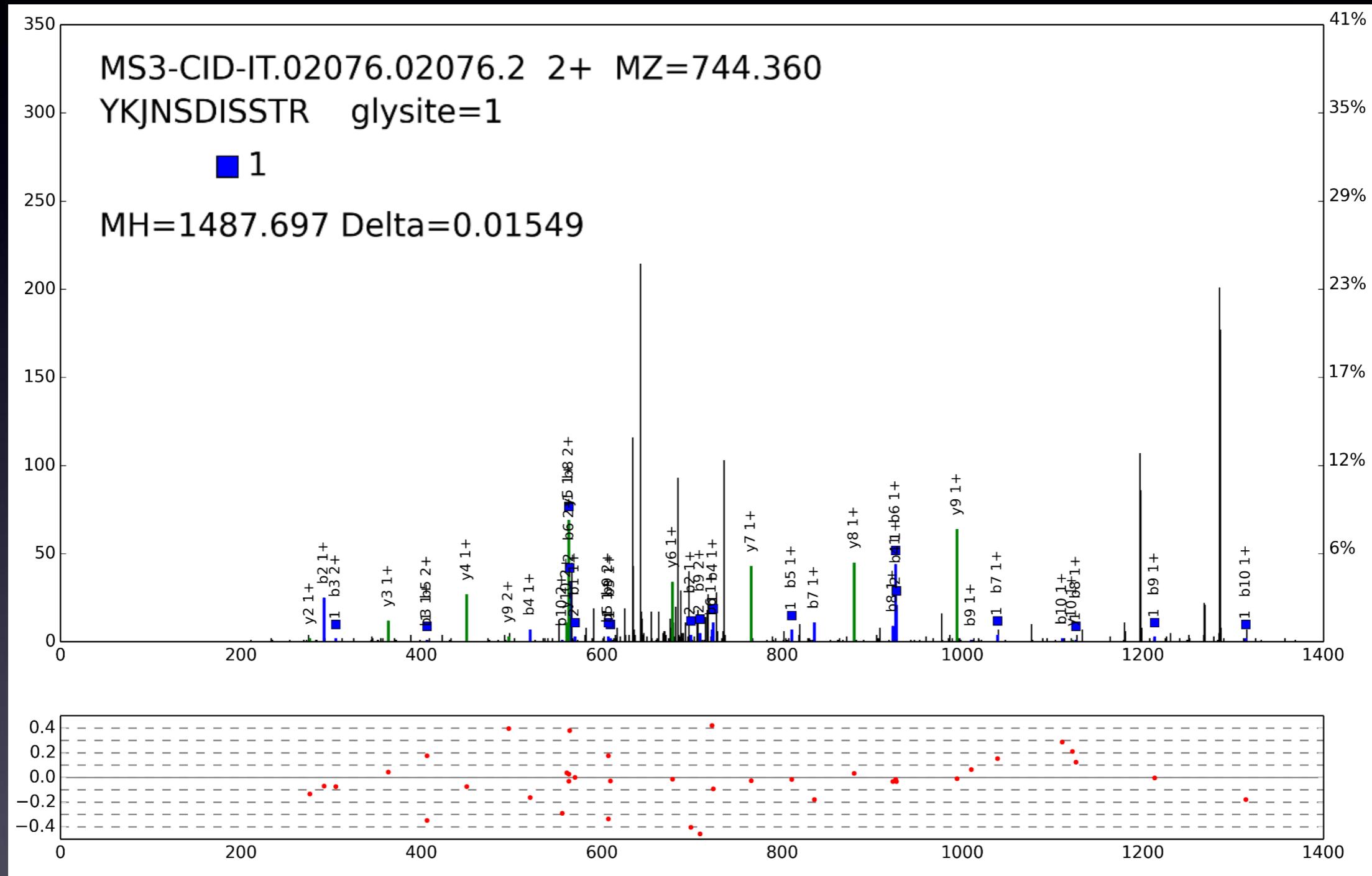
Long term goal:
Interpretation of intact glycopeptide in complex sample

- Possible strategies

- Our pipeline

- ***Challenges***

Application of our pipeline to complex sample: Potential challenge

- MS/MS spectral processing

Protein-database free algorithm should work fine

- MS/MS spectral collection

We intentionally split MS/MS and MS3 spectral collection into two MS runs, therefore the balance between quality and quantity should work well

- MS3 spectral processing

Conventional protein database searching of tryptic peptide

- MS3 spectral collection

Is the sensitivity of MS3 enough?

Application of our pipeline to complex sample: Challenge in MS3

- For MS/MS, the latest instrument shows almost zero difference in terms of sensitivity between IT and OT.

- However, for MS3, it is completely different.

DDA MS3 from HCD-MS/MS

IT as detector → Results shown today

OT as detector → Nothing
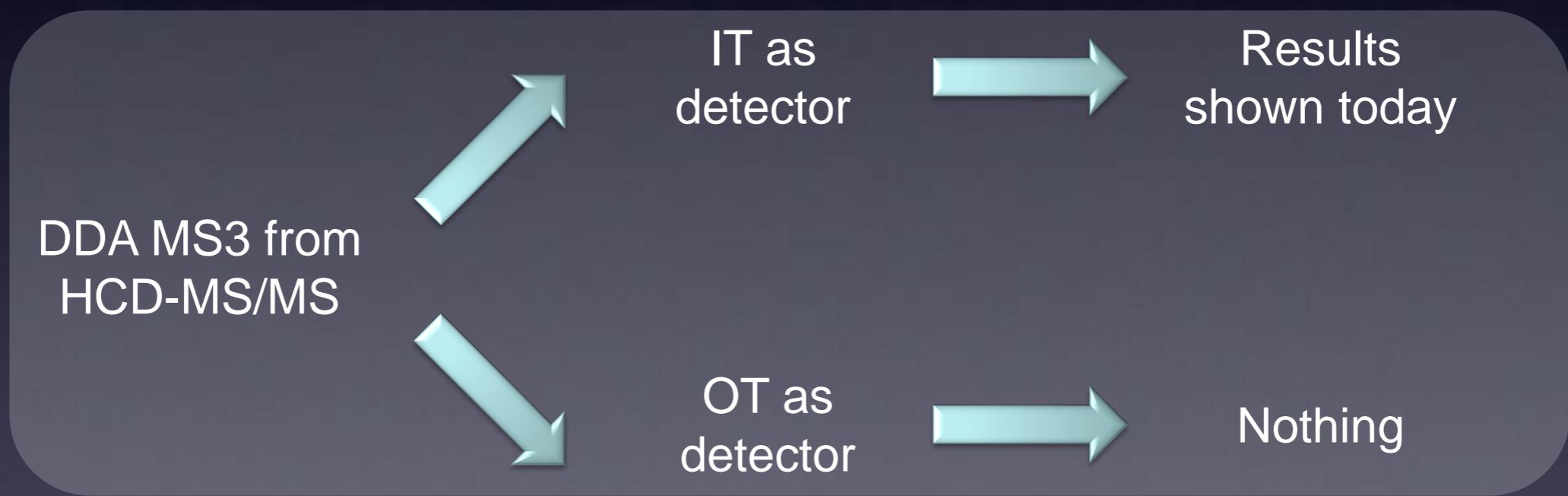
AGC target: maximum
Accumulation time: 500ms

Application of our pipeline to complex sample:
Top 500 MS3 TIC - IT vs OT
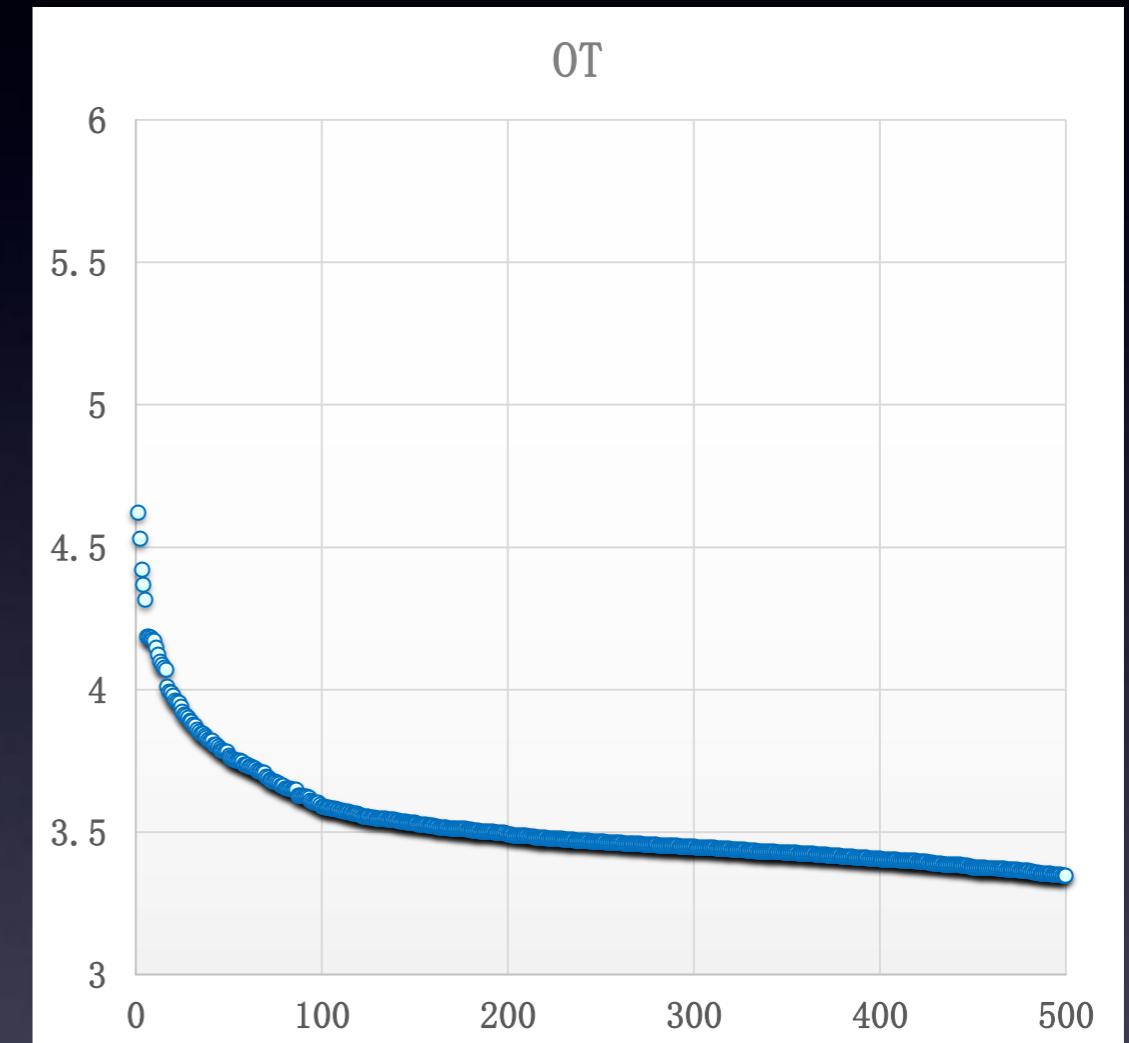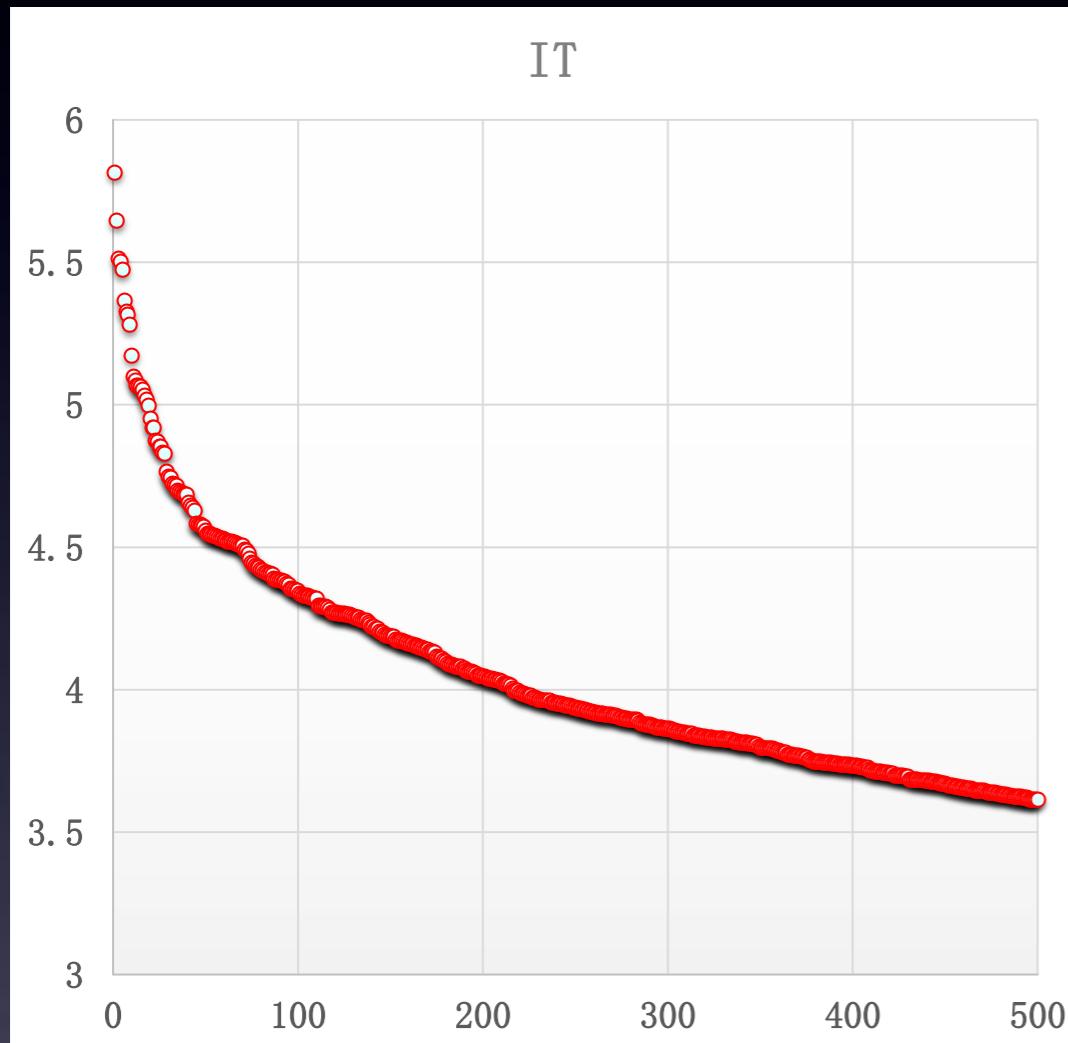
Conclusion

- We have developed Integrated computational pipeline for the complete sequencing of glycopeptide using HCD/CID-MS/MS and MS3

- We will try to interpret glycopeptide in complex sample after some fine tuning in MS3 spectral collection

Acknowledgement