

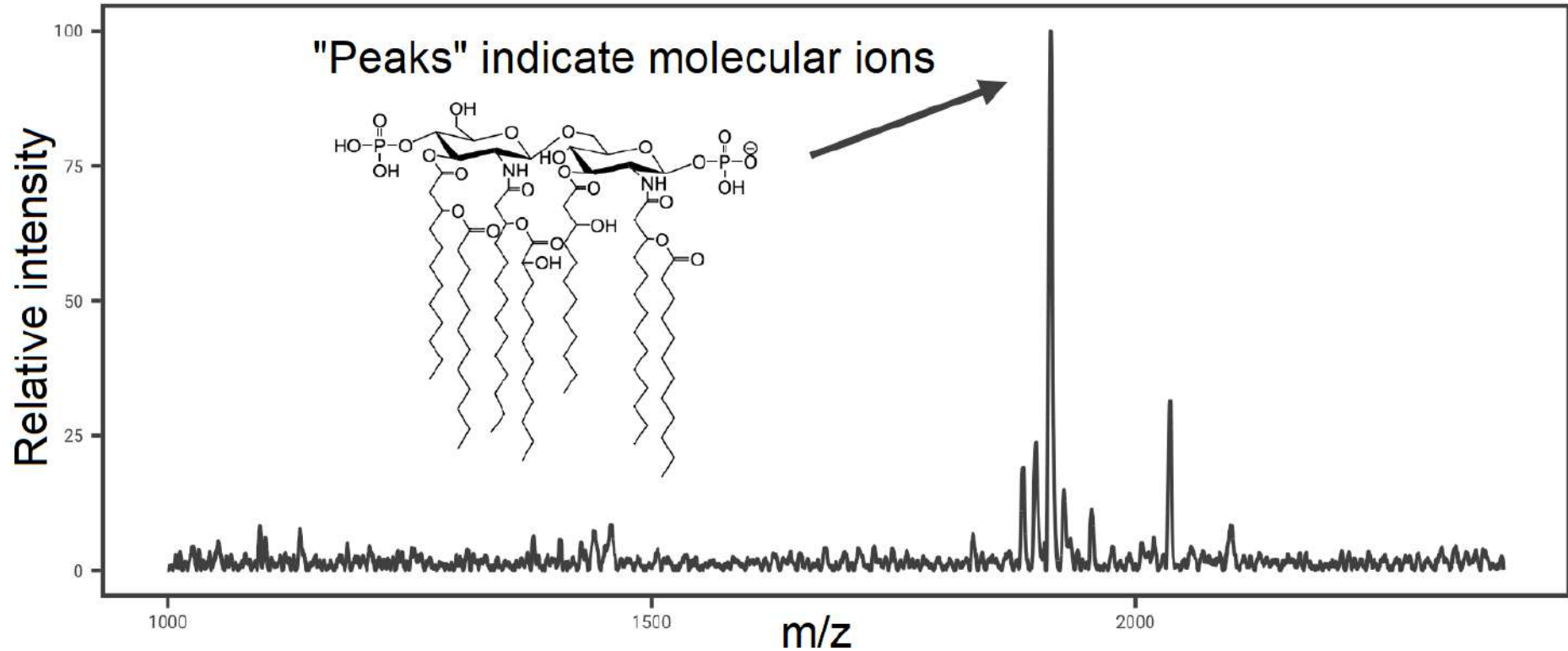


DIAmeter : Matching peptides to data-independent acquisition mass spectrometry data

[Yang Lu](#), Jeff Bilmes, Ricard Rodriguez-Mias, Judit Villen, William Stafford Noble

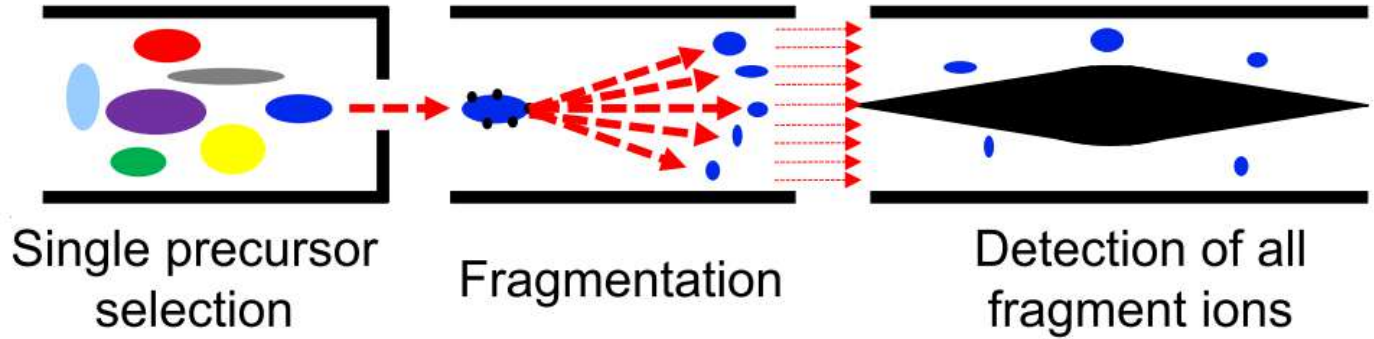


The mass spectrum is a record of the detected ions

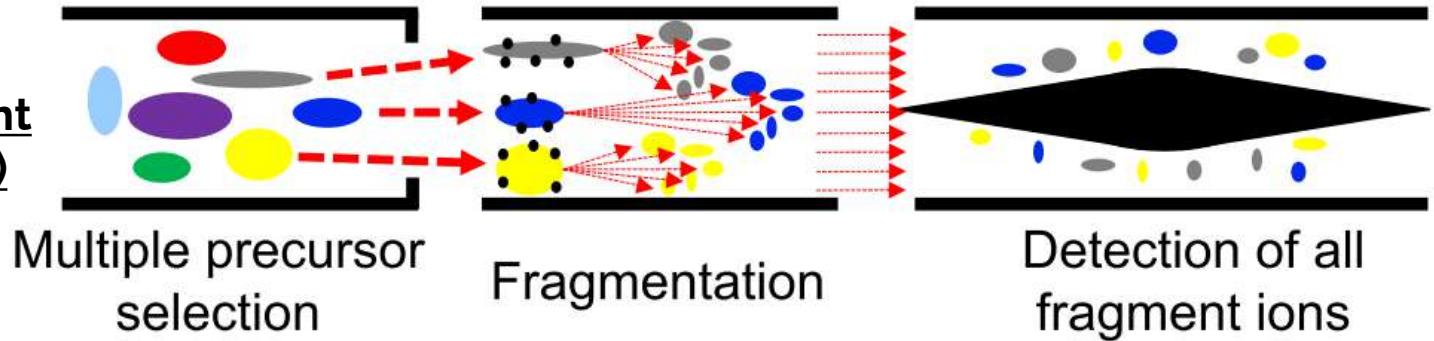


The mass spectrum is obtained by isolating and fragmenting peptides

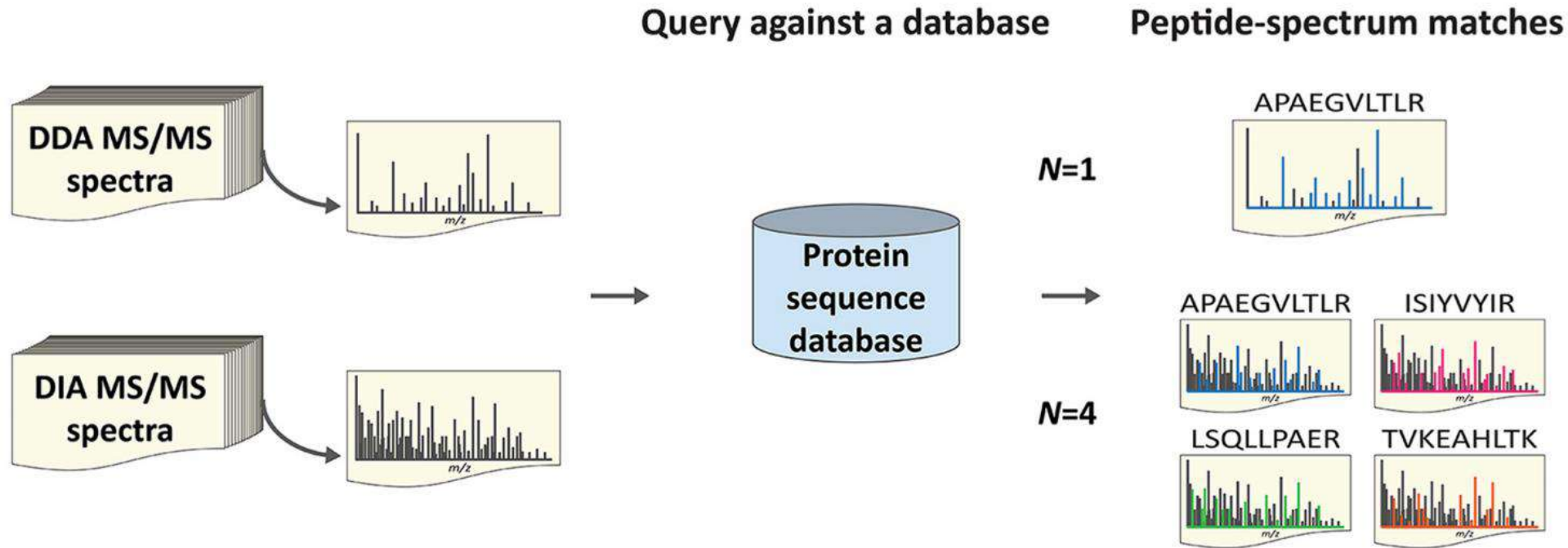
Data-dependent acquisition (DDA)



Data-independent acquisition (DIA)



Standard proteomics database search process is typically designed for DDA data

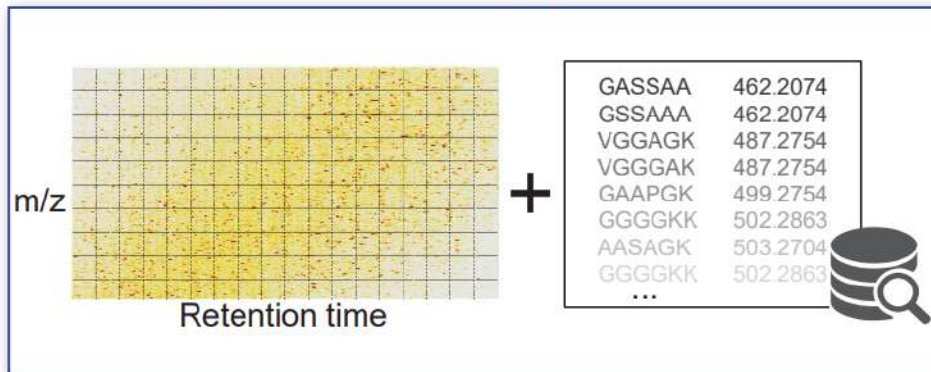


There is demand for new methods that overcome limitations of existing methods

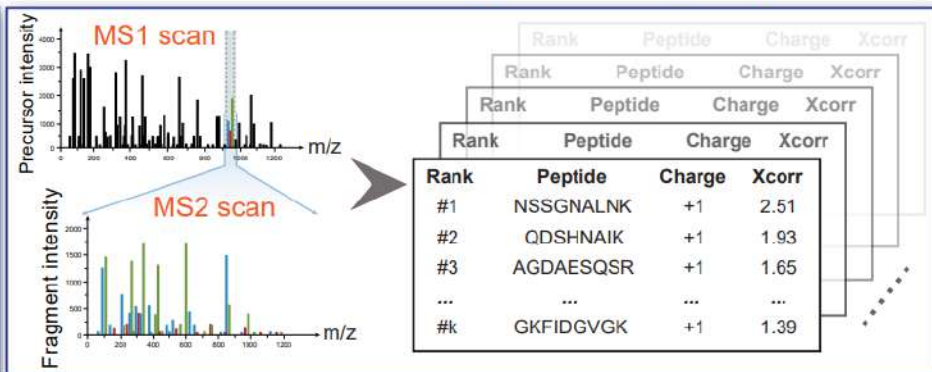
	DIAmeter	DIA-Umpire	PECAN	Prosit + EncyclopeDIA
Support wide isolation windows	✓	✓	✗	✗
Detects PTMs	✓	✓	✓	✗
Instrument independent	✓	✓	✗	✗
Detects peptides with undetectable precursor	✓	✗	✓	✓

DIAMeter uses a standard DDA search engine with an extra PSM filtering step

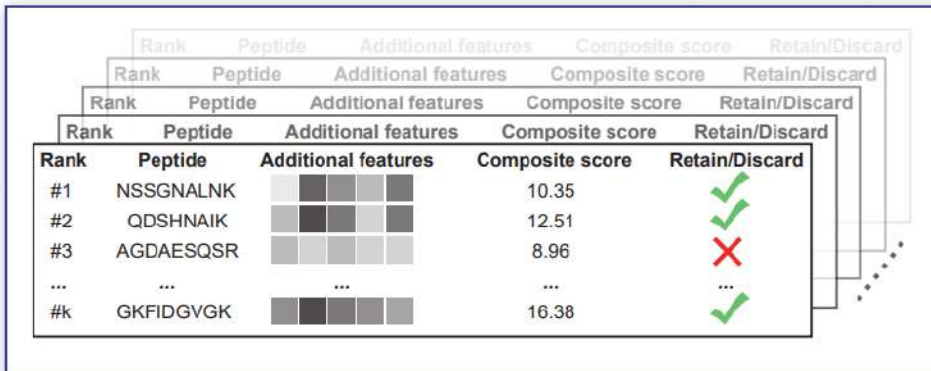
1. Input: DIA data and a peptide database



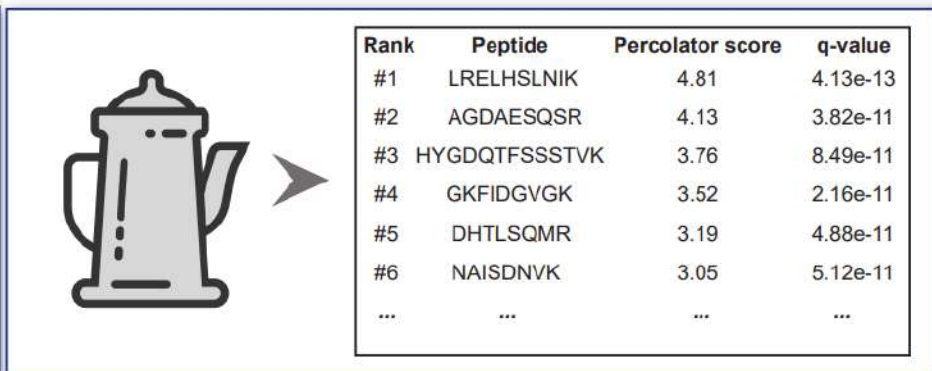
2. Tide search identifies multiple matches per spectrum



3. Matches are reduced by using PSM features



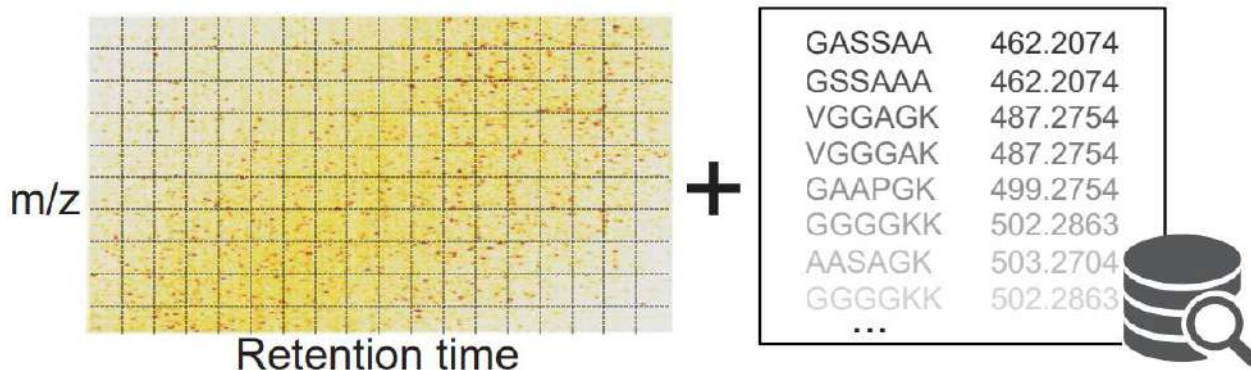
4. DIA-tailored Percolator ranks peptides by q-value



DIAMeter searches spectra by using a peptide sequence database

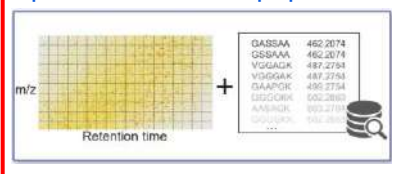
Input:

- ❑ DIA data
- ❑ A peptide database

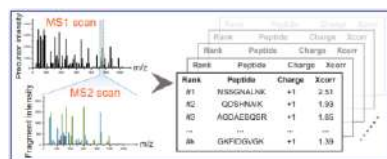


At a high level:

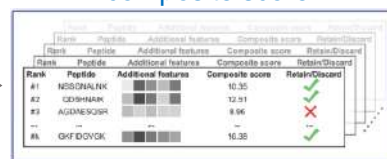
Input: DIA data and peptide DB



Tide search identifies PSMs



PSMs are filtered by a composite score



Percolator prioritizes peptides



DIAMeter uses Tide to identify matches per spectrum

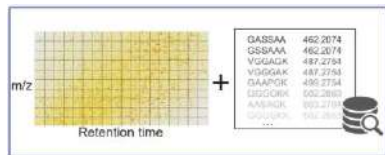
Tide search for PSMs:

- ❑ Use **XCorr** as the primary score
- ❑ Identify top-k (k=5) matches for each spectrum and each charge state

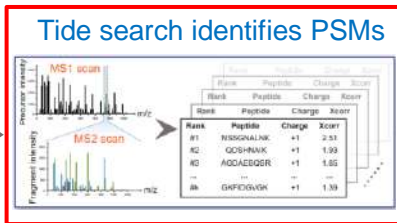
Rank	Peptide	Charge	Xcorr
Rank	Peptide	Charge	Xcorr
Rank	Peptide	Charge	Xcorr
Rank	Peptide	Charge	Xcorr
#1	NSSGNALNK	+1	2.51
#2	QDSHNAIK	+1	1.93
#3	AGDAESQSR	+1	1.65
...
#k	GKFIDGVGK	+1	1.39

At a high level:

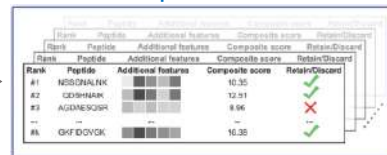
Input: DIA data and peptide DB



Tide search identifies PSMs



PSMs are filtered by a composite score



Percolator prioritizes peptides



DIAMeter filters the PSMs by a composite score

The composite score is defined as the **weighted sum** of **PSM features**:

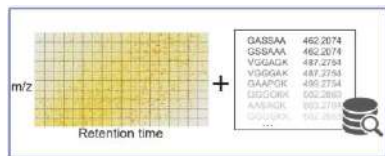
- ☐ XCorr (with Tailor calibration)
- ☐ Precursor intensity rank
- ☐ Fragment matching p-value
- ☐ Retention time score
- ☐ Precursor and fragment coelution

Rank	Peptide	Additional features	Composite score	Retain/Discard
#1	NSSGNALNK		10.35	✓
#2	QDSHNAIK		12.51	✓
#3	AGDAESQSR		8.96	✗
...
#k	GKFIDGVGK		16.38	✓

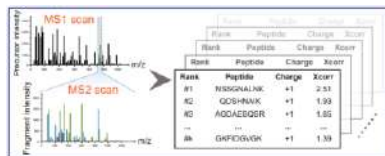
Sulimov et al. Journal of Proteome Research (2020)

At a high level:

Input: DIA data and peptide DB



Tide search identifies PSMs



PSMs are filtered by a composite score

Rank	Peptide	Additional features	Composite score	Retain/Discard
#1	NSSGNALNK		10.35	✓
#2	QDSHNAIK		12.51	✓
#3	AGDAESQSR		8.96	✗
...
#k	GKFIDGVGK		16.38	✓

Percolator prioritizes peptides



Rank	Peptide	Percolator score	q-value
#1	URELHILNK	4.81	4.13e-10
#2	AGDAESQSR	4.10	3.82e-11
#3	HYDDQTFSSSTYK	3.76	8.43e-11
#4	GKFIDGVGK	3.52	2.19e-11
#5	DHITLSQMP	3.19	4.89e-11
#6	NAISDNVK	3.05	5.12e-11
...

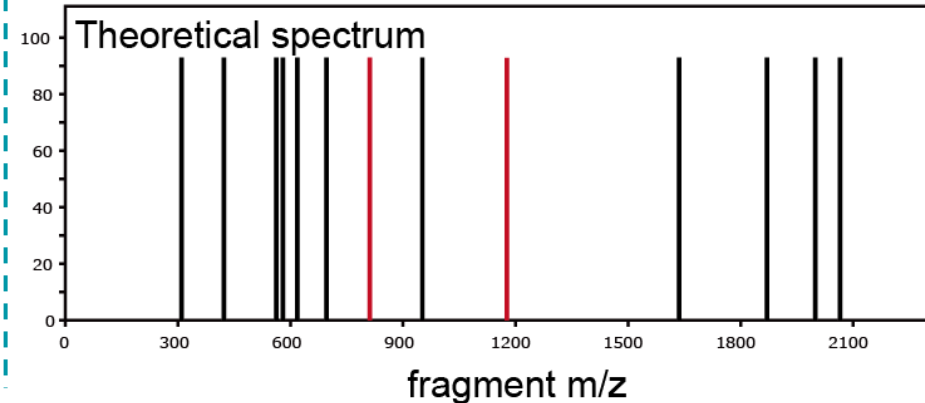
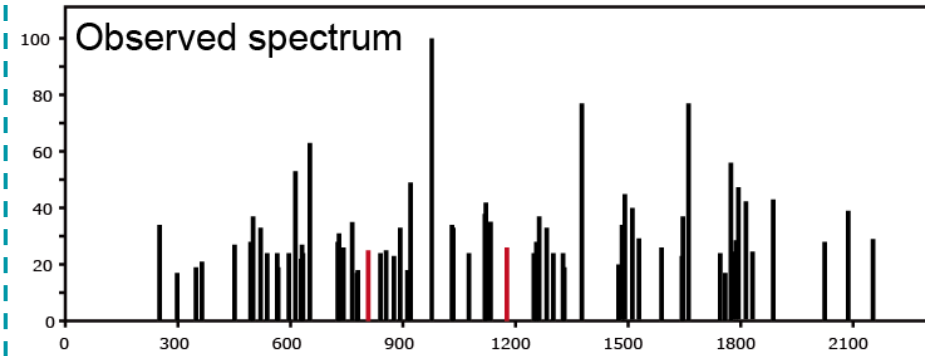
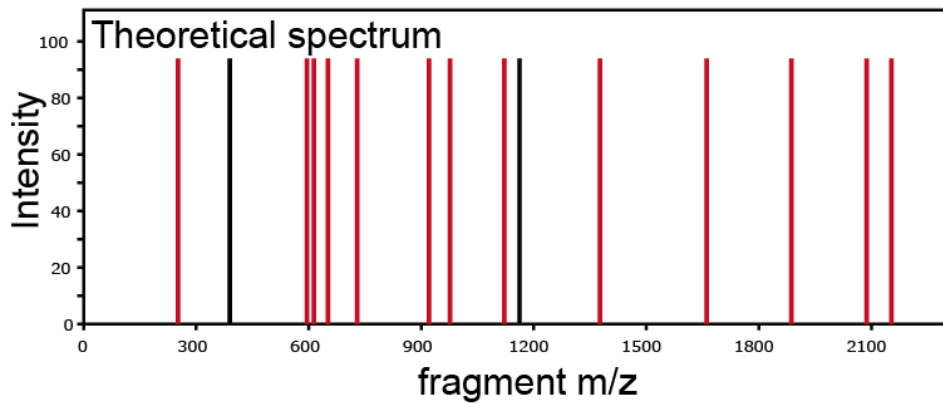
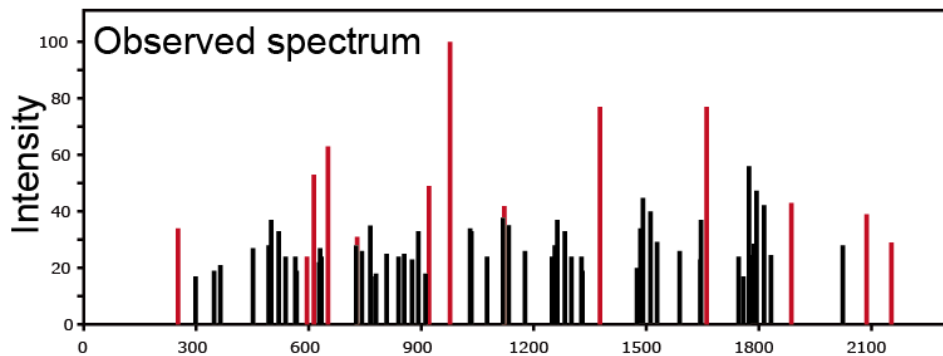
XCorr score measures the fragment-level matching evidence



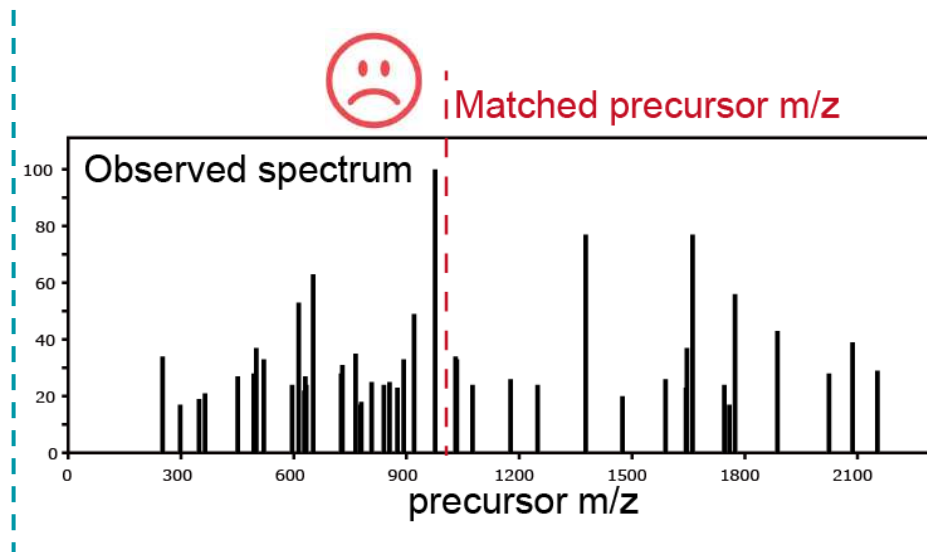
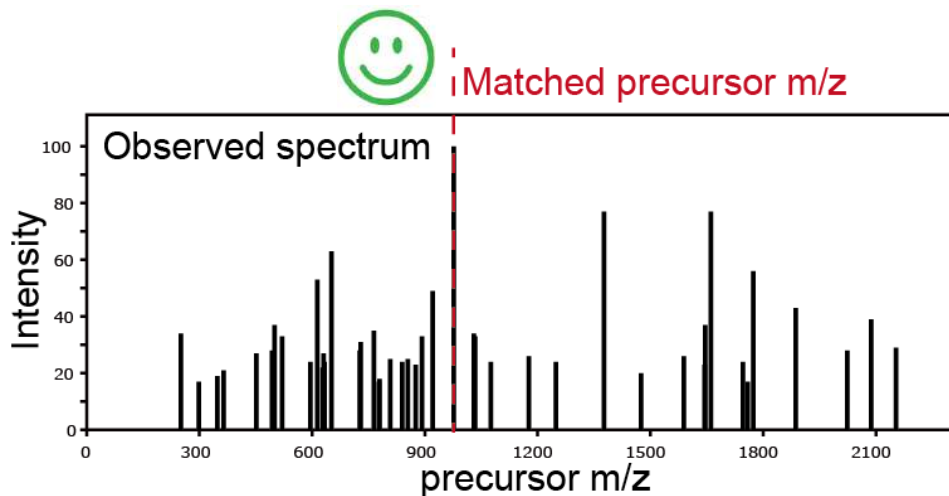
Unmatched peak



Matched peak



Precursor intensity measures the precursor-level matching evidence



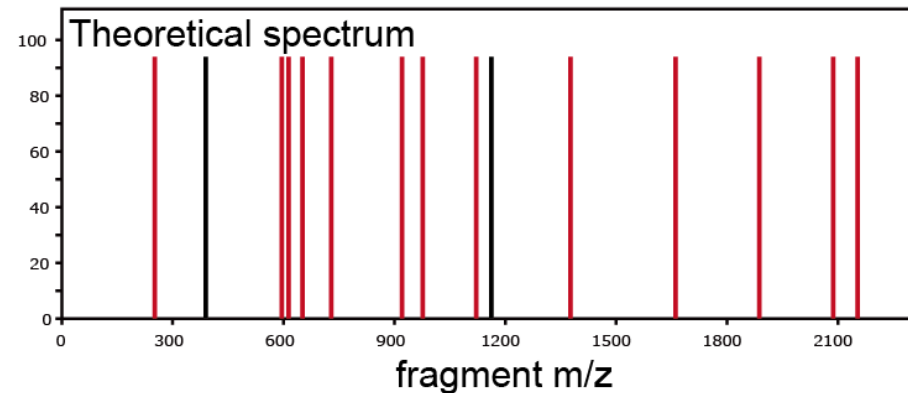
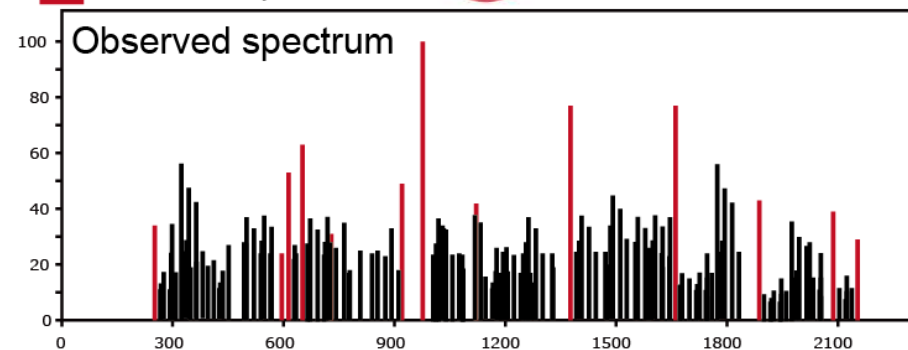
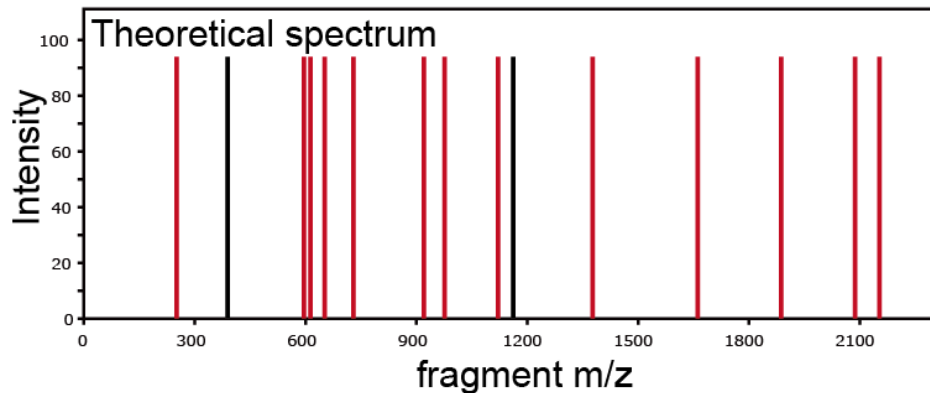
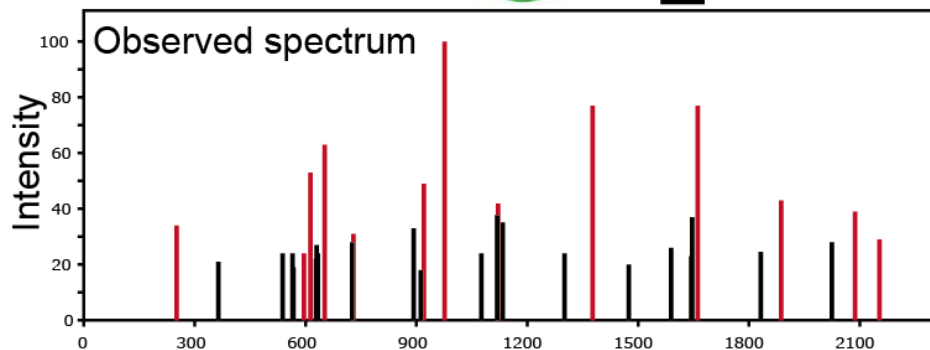
Fragment matching p-value measures how likely the theoretical fragments are to match by random chance.



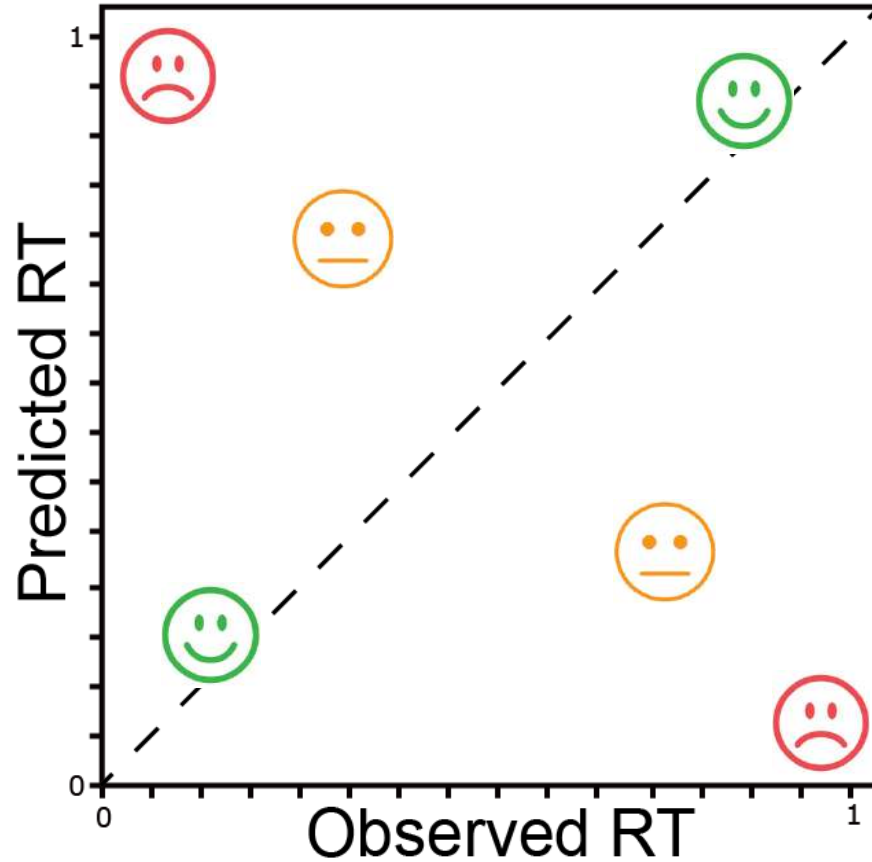
Unmatched peak



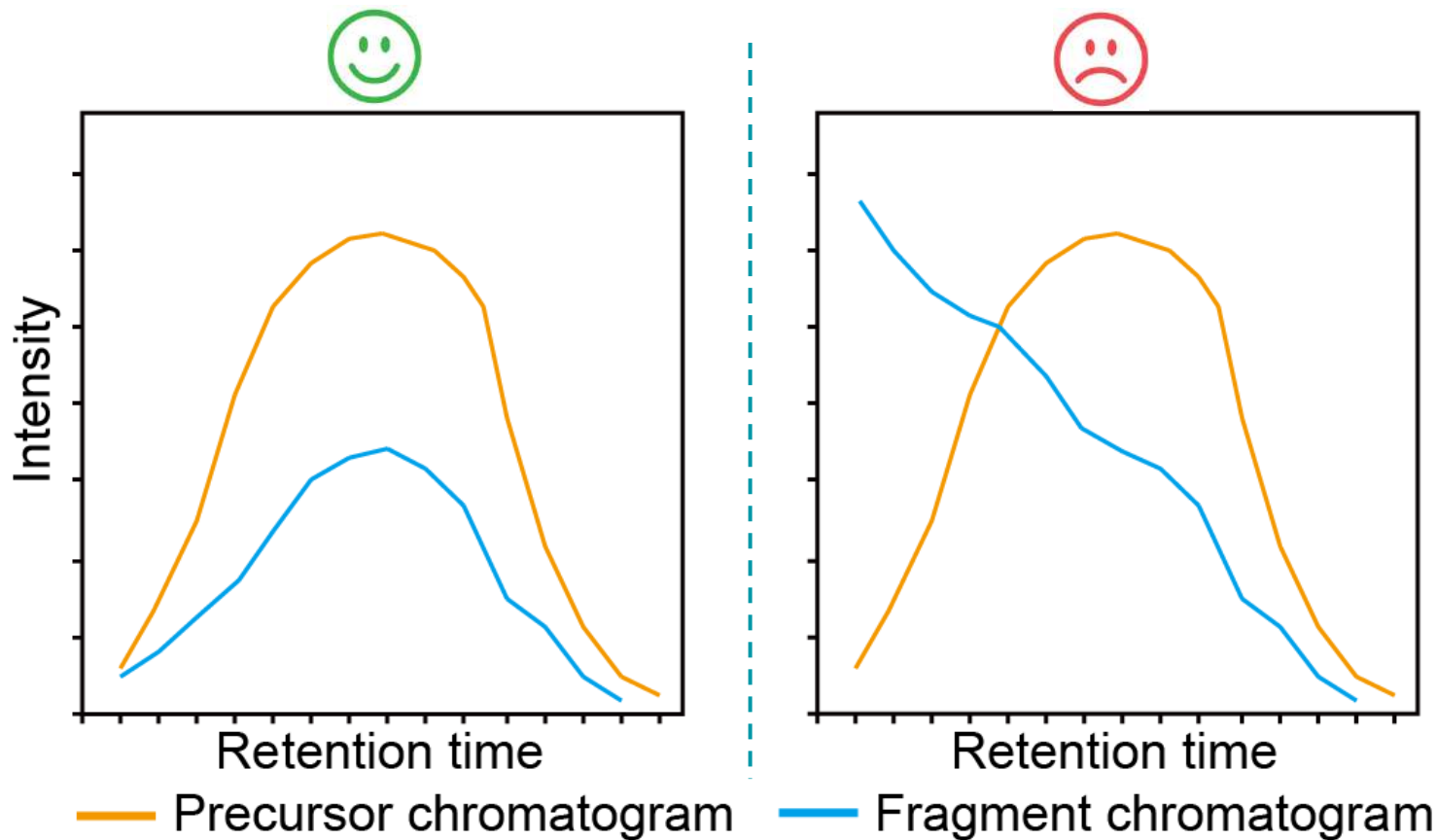
Matched peak



Retention time score measures the temporal matching evidence



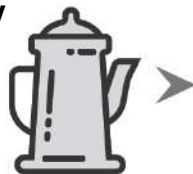
Precursor and fragment coelution measures the chromatogram-level matching evidence



DIA-tailored Percolator ranks peptides by q-value

Percolator learns to re-rank these PSMs:

- ❑ The **modified** Percolator assigns all PSMs related to the same target-decoy pair to the same cross-validation split
- ❑ DIAMeter selects the top-scoring PSM per target-decoy pair
- ❑ DIAMeter calculates its own peptide-level FDR estimates.

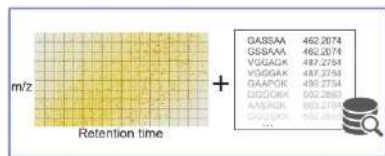


Rank	Peptide	Percolator score	q-value
#1	LRELHSLNIK	4.81	4.13e-13
#2	AGDAESQSR	4.13	3.82e-11
#3	HYGDQTFSSSTVK	3.76	8.49e-11
#4	GKFIDGVGK	3.52	2.16e-11
#5	DHTLSQMR	3.19	4.88e-11
#6	NAISDNVK	3.05	5.12e-11
...

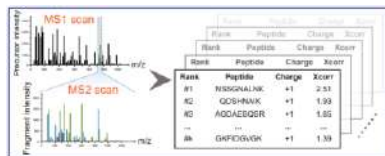
Kall et al. Nature Methods (2007)

At a high level:

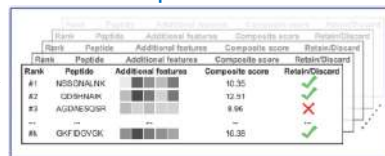
Input: DIA data and peptide DB



Tide search identifies PSMs



PSMs are filtered by a composite score

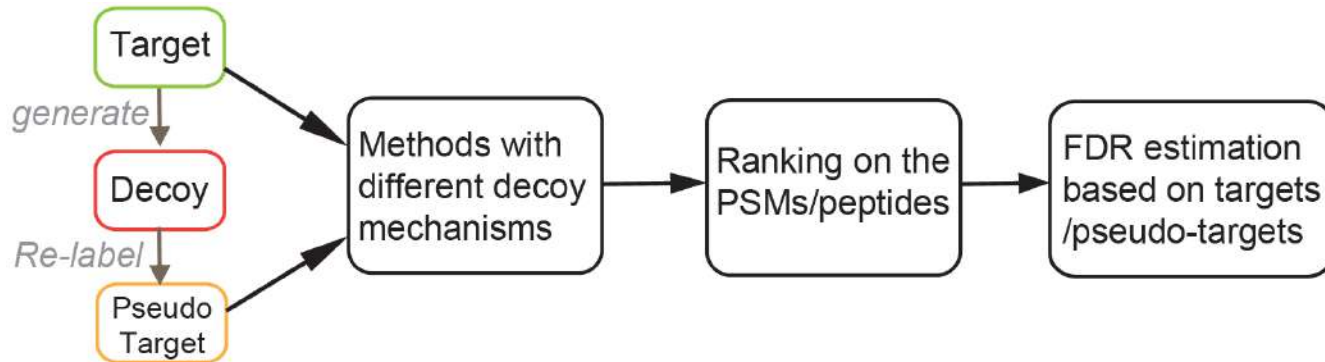


Percolator prioritizes peptides



Evaluation: Consistent FDR estimation across methods by using an entrapment strategy

- ❑ Methods are agnostic to pseudo targets
- ❑ Pseudo targets leads to loss of statistical power, only for evaluation purposes



DIAMeter is evaluated on datasets across diverse settings

Yeast dataset with enriched methionine oxidation (OxMetYeast)

- ❑ Orbitrap instrument
- ❑ Acquired by using different isolation window sizes

Multispecies mixture dataset (LFQbench)

- ❑ TripTOF 5600/6600 instrument
- ❑ Acquired by using fixed/variable isolation window

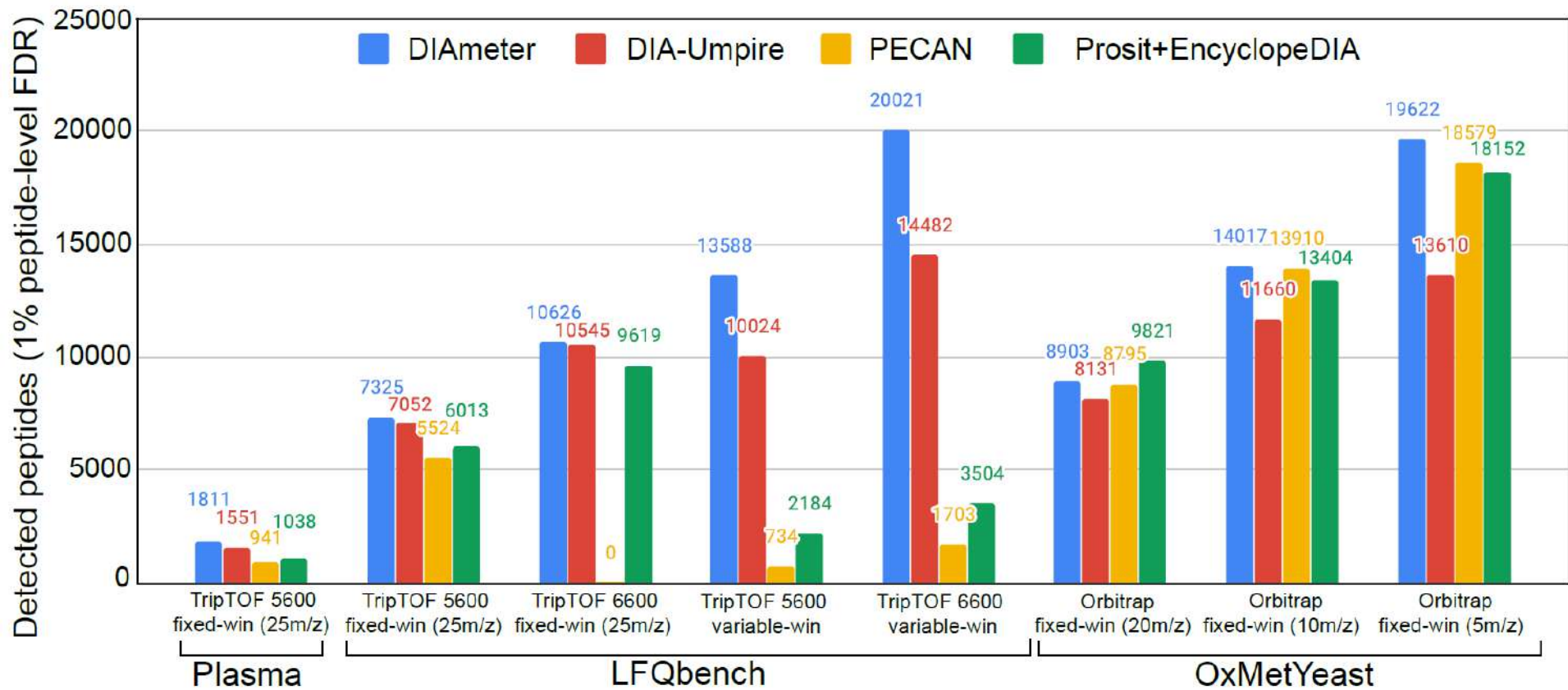
Navarro et al. Nature Biotechnology (2016)

Human blood plasma dataset (Plasma)

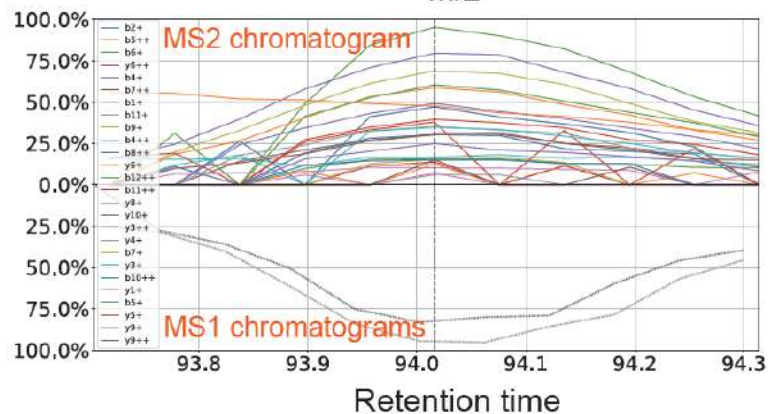
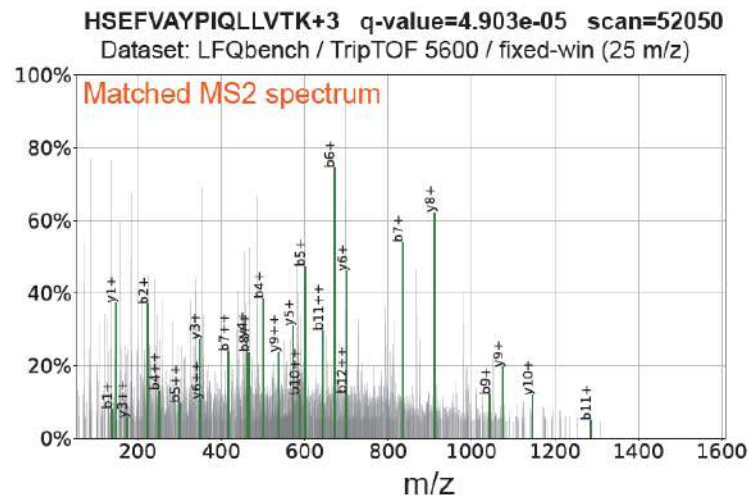
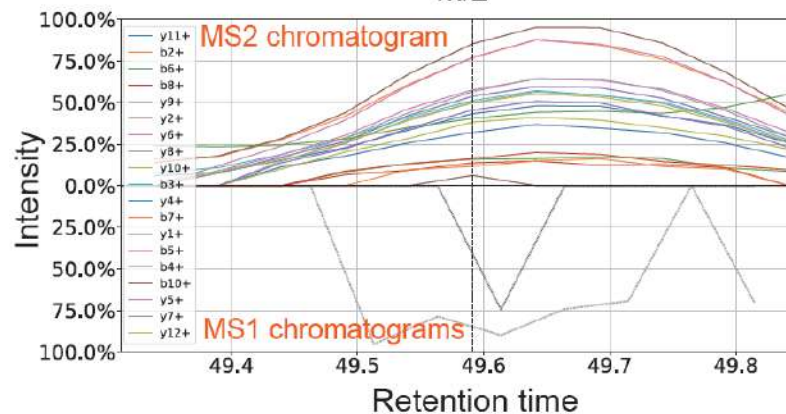
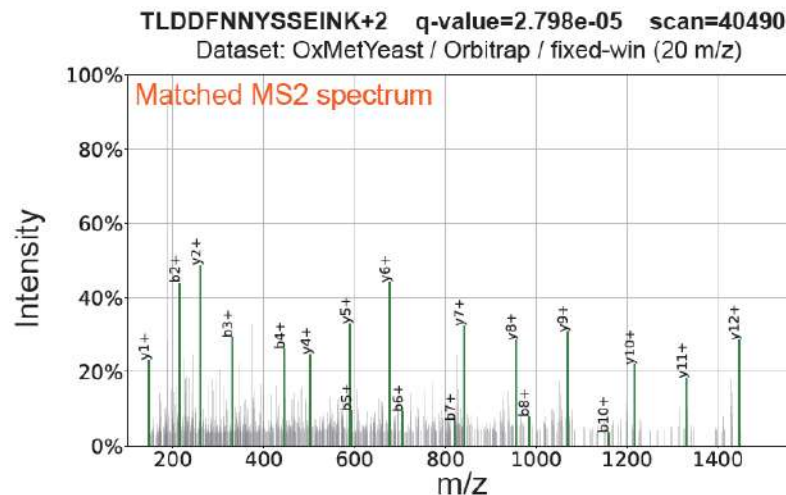
- ❑ TripTOF 5600 instrument
- ❑ Acquired by using variable isolation window
- ❑ High dynamic range

Rosenberger et al. Nature Biotechnology (2017)

DIAMeter confidently detects more peptides at 1% FDR



DIAMeter can detect peptides with undetectable MS1 signal

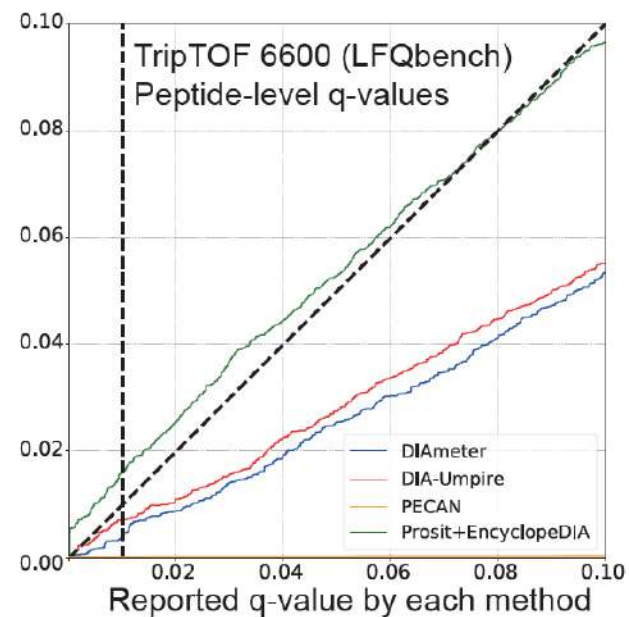
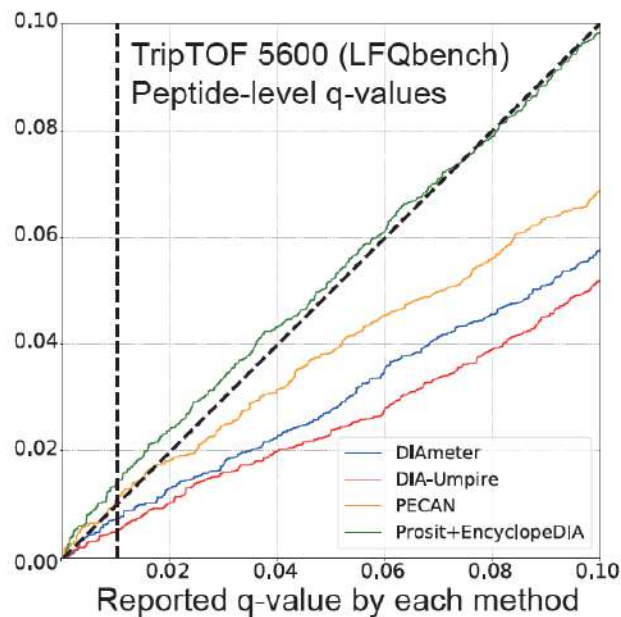
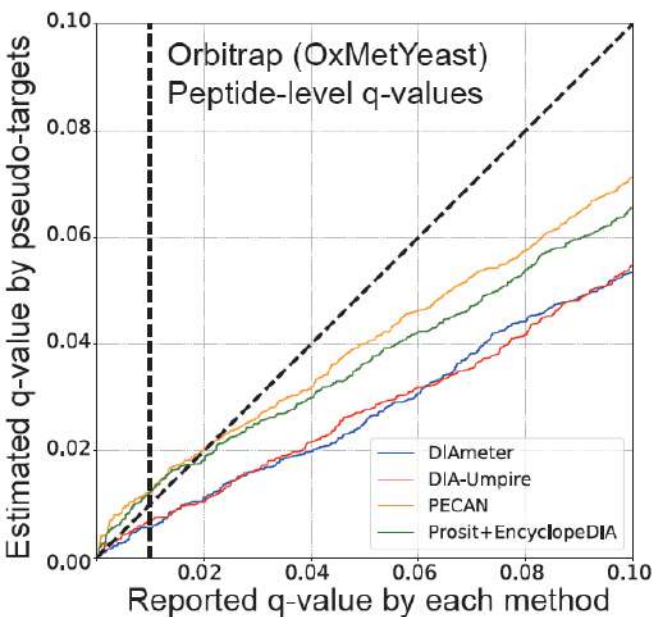


— Precursor intensity (monoisotope)

- - - Precursor intensity (M+1 isotope)

- - - Precursor intensity (M+2 isotope)

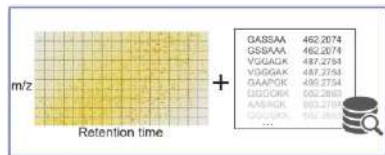
DIAMeter accurately controls the FDR



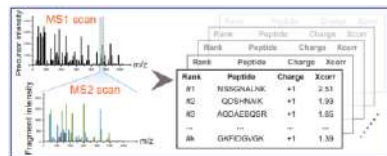
Conclusions

- ❑ DIAMeter is more sensitive than other methods, without relying on spectral library
- ❑ DIAMeter is robust in diverse settings
- ❑ DIAMeter can detect peptides with undetectable MS1 signal
- ❑ Availability: <http://crux.ms/>

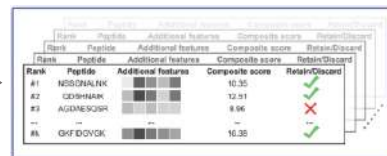
Input: DIA data and peptide DB



Tide search identifies PSMs



PSMs are filtered by extra features



Percolator prioritizes peptides



Acknowledgements

- Noble lab members:

