

Practical Consideration for Acquisition and Analysis of DIA/SWATH-MS dataset

Mukul K. Midha, Ph.D.

03/23/2019

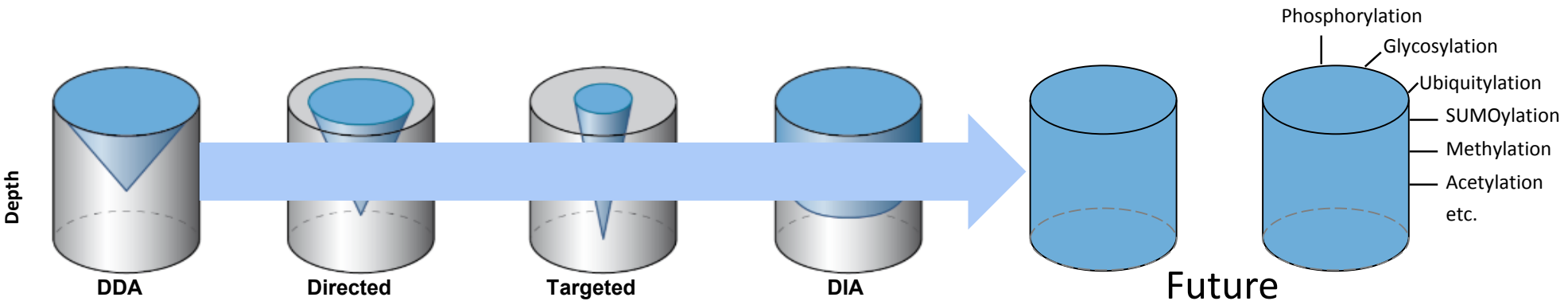
Moritz group

Institute for Systems Biology, Seattle

Outline

- DIA/SWATH-MS basic concepts
- DIA/SWATH-MS objectives and workflows
- DIA/SWATH-MS analysis tools

Mass spectrometry based proteomics



Shotgun/
discovery

Shotgun
with
inclusion
lists

Selected
Reaction
Monitoring

SWATH

Comprehensive
Sensitive
Reproducible

& Protein
Modifications

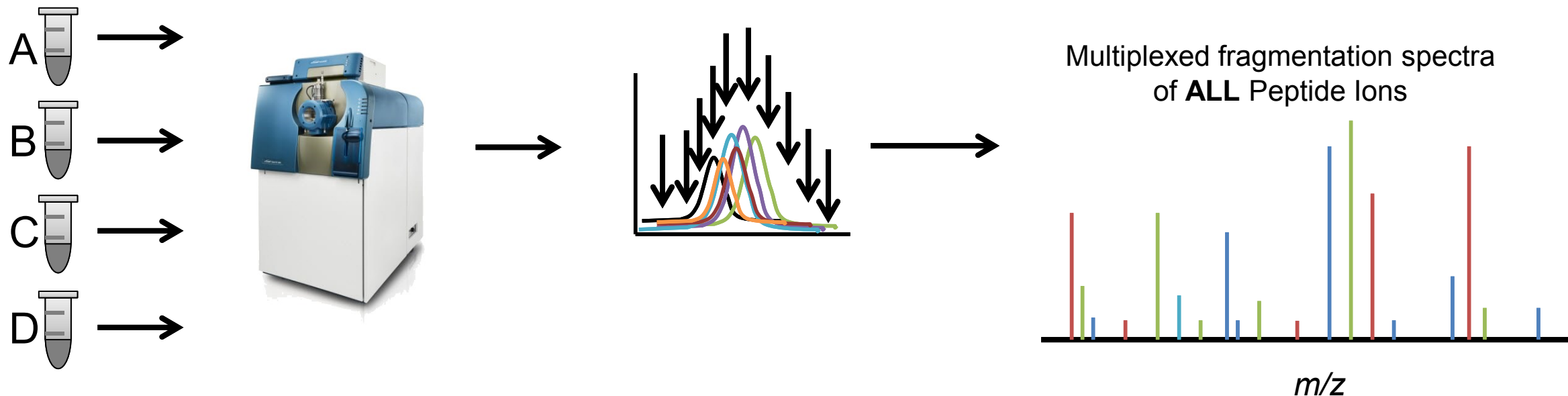
Adapted from Ariel Bensimon et al, Annu. Rev. Biochem. 2012, 81, 379.

- protein composition of a sample
- thousands of peptides
- stochastic nature of precursor selection

- lower limit of detection, wider dynamic range
- increased reproducibility
- limited number of peptides



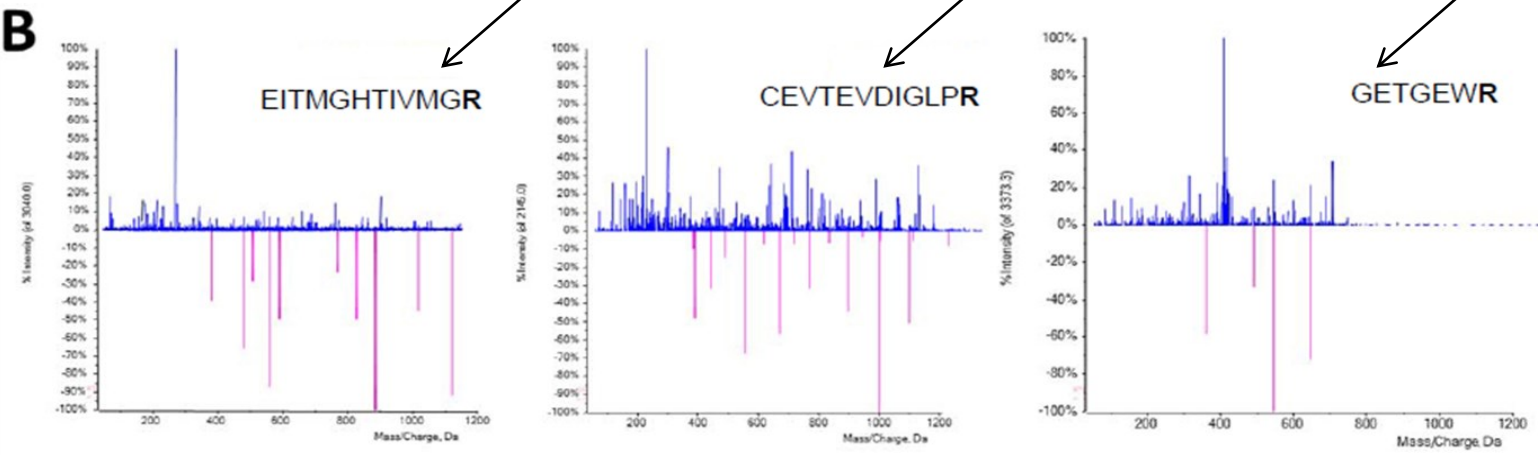
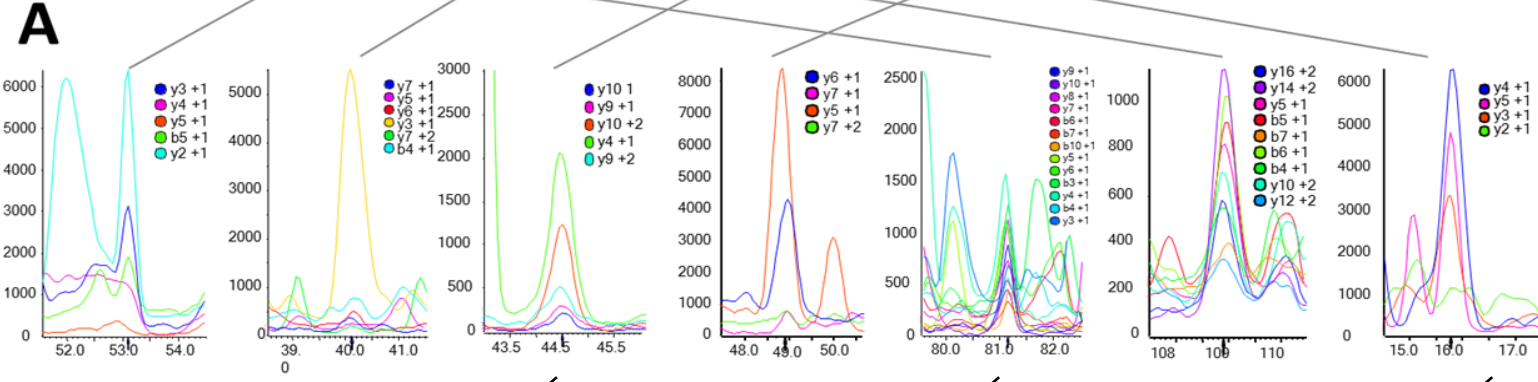
SWATH-MS- Discovery Proteomics



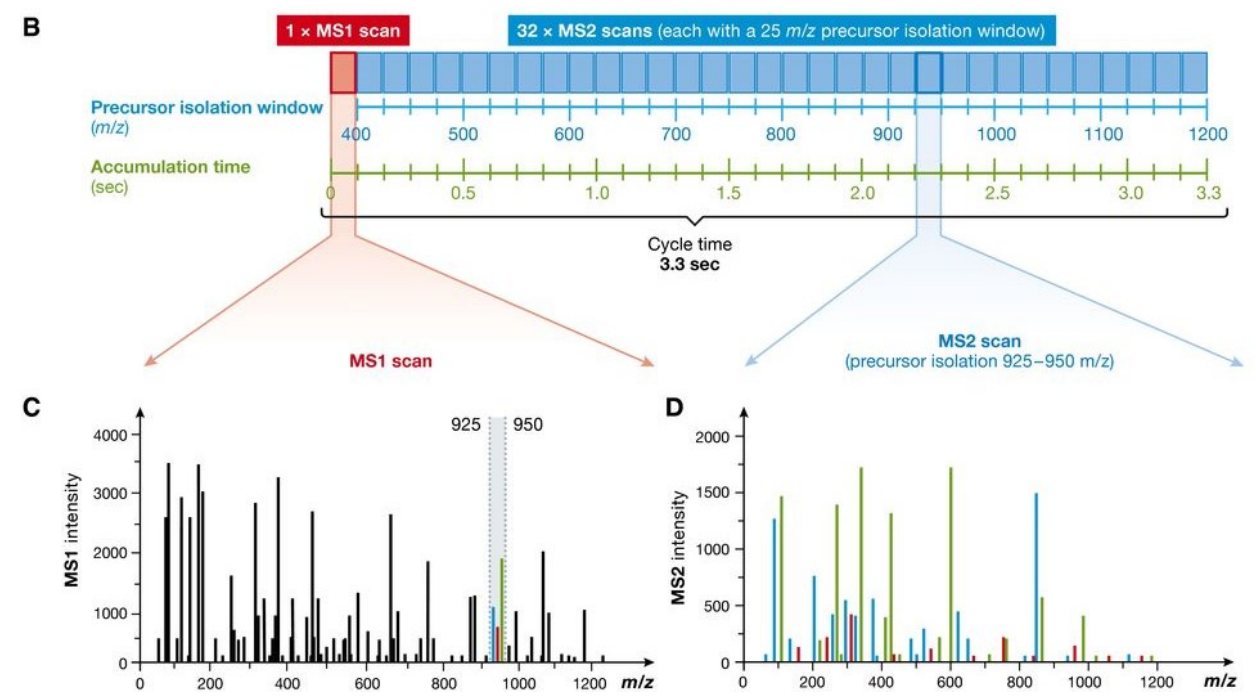
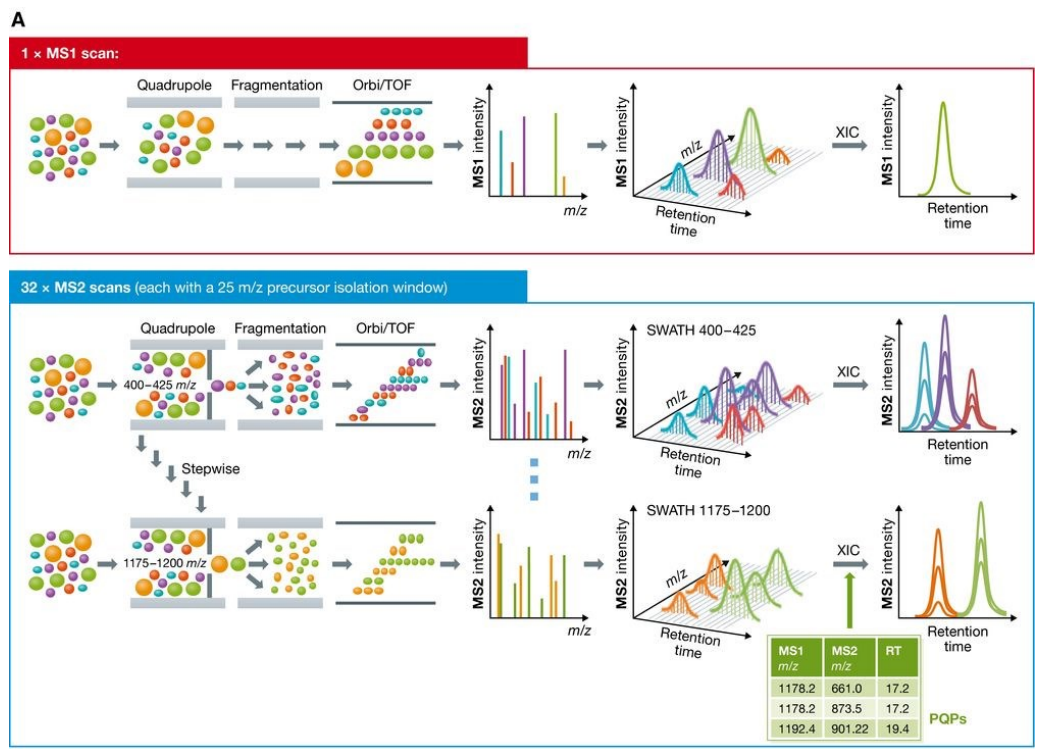
- Each sample is directly analyzed without fractionation.
- Selection of peptides is multiplexed.
- Results are multiplexed spectra to be identified by known spectral library
- Next stage is computational analysis.

SWATH-MS targeted data analysis

MVGLIWAQATSGVIGR **GGDIPWR** **LPEDQAHFR** **EITMGHTIVMGR** R **TWDSLPAK**
 VRPLPGR R NVVLSR QADFMASGAEVVGSLEEALTSPE^TWVIGGGQVYALALPYATR
CEVTEVDIGLPR **EAGDALAPVLD**^T**ETWR** **GETGEWR** FSR SGLR YR LYSYHR S



Principle of sequentially windowed acquisition in DIA/SWATH-MS



SWATH-MS/DIA experiment setup....So much to choose from!!

Instruments.....

Nano flow or
Micro flow?

LC separation
considerations?

Sample
datasets...

Library free
approach

Spectral ion
Libraries

Fixed or Variable
windows?

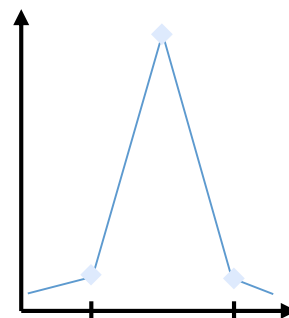
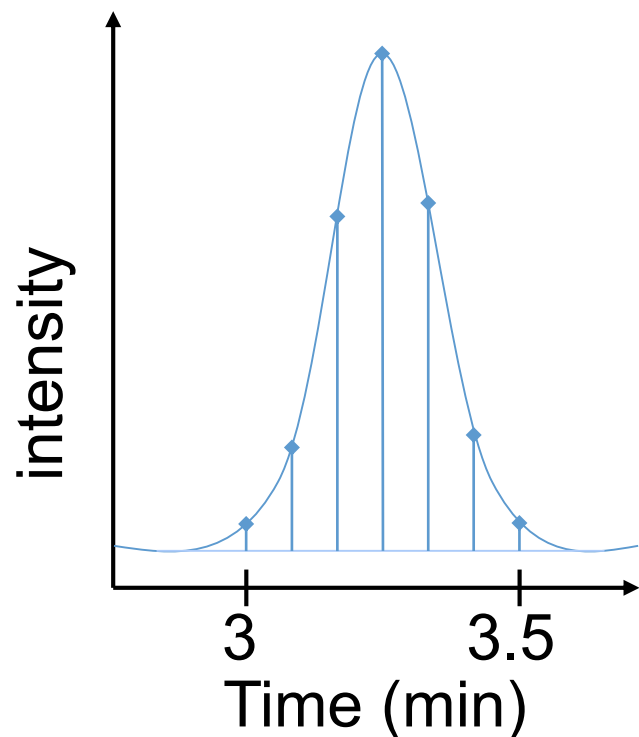
Software, software,
software.....

What are the critical acquisition attributes for DIA/SWATH-MS?

- ✓ High resolution MS/MS
- ✓ Cycle Time
- ✓ Q1 Isolation Windows
- ✓ Dynamic Range

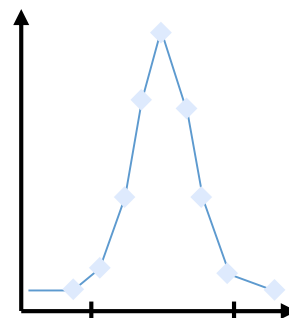


DIA Acquisition- Quantitation

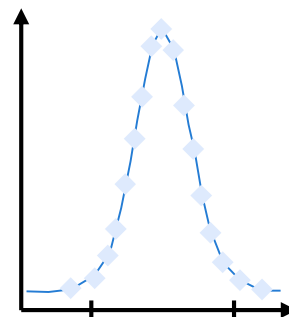


Data Points Per Peak (DPPP)

<7 DPPP = under sampling



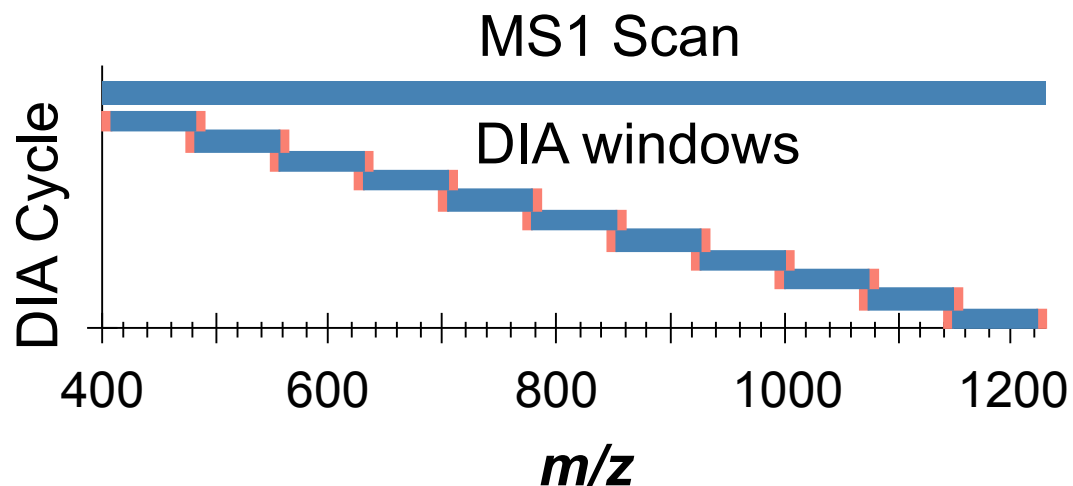
7-10 DPPP = optimal sampling



>10 DPPP = over sampling

ex. 30 s peak width at base
3 s cycle will collect 10 DPPP
<3 s will over sample
>3 s will under sample

DIA/SWATH-MS precursor isolation windows

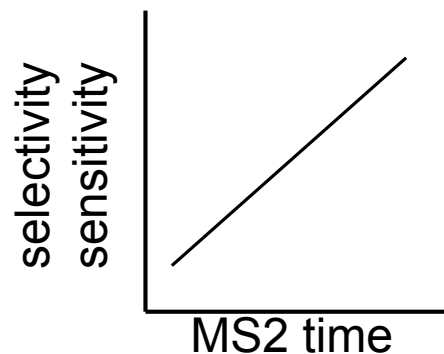
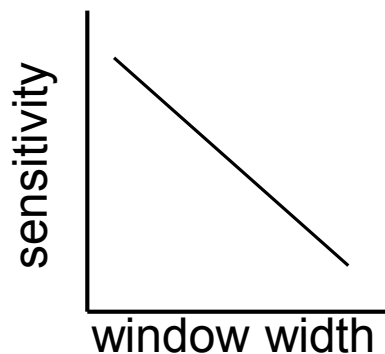


Window Strategies

- sequential segments
 - w/ or w/o overlap
 - static or variable width

windows x MS2 acquisition time = cycle time

42 (20 m/z width, 400-1200 m/z, 1 Da overlap) x 60 ms = 3.6 s

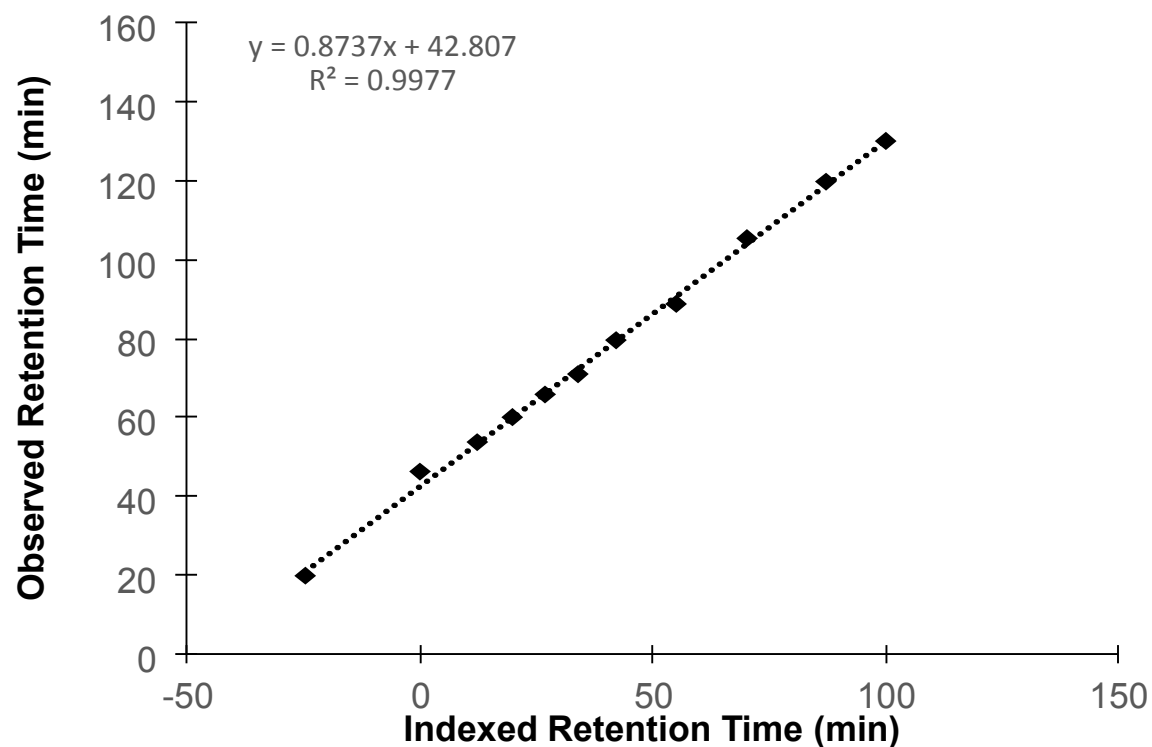


Choices are dependent on chromatography, application and platform.

DIA data analysis – retention time normalization

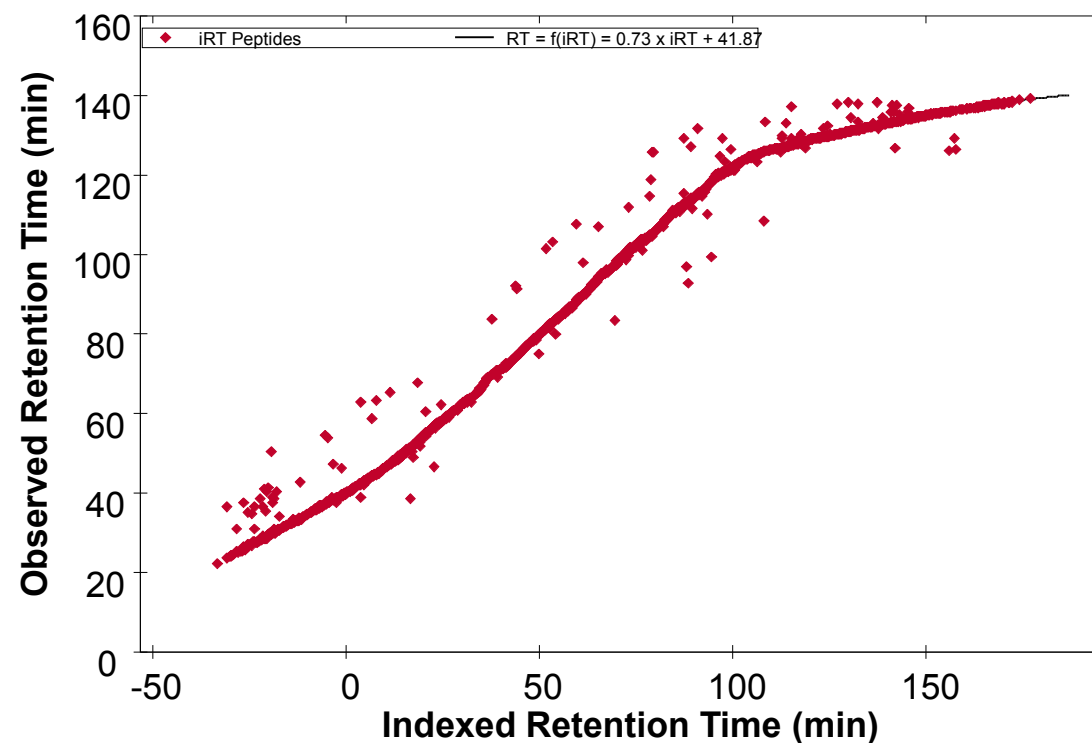
■ iRT spike

- set of non-endogenous peptides
- Used to convert to iRT scale (linear)

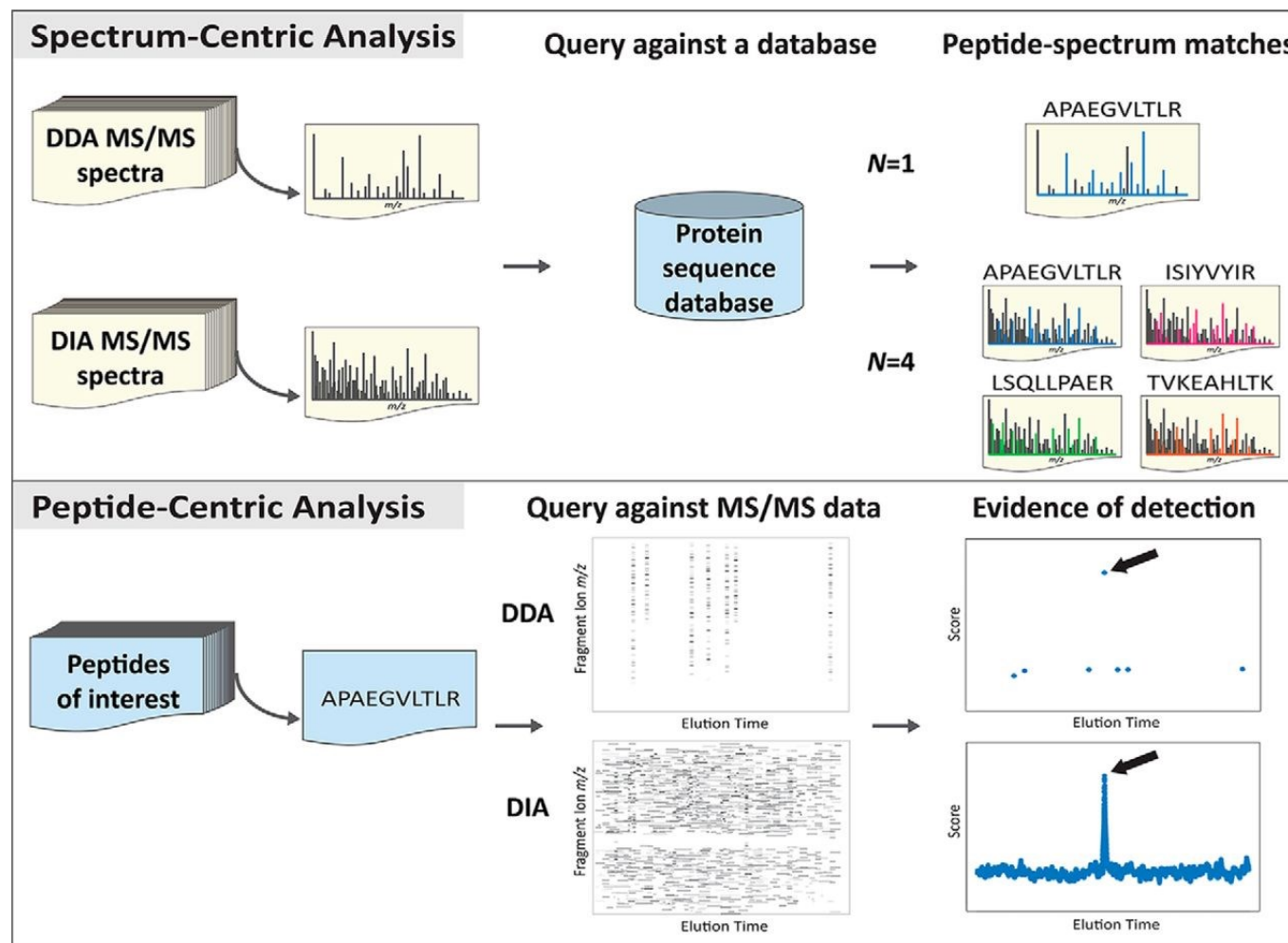


■ Observed peptides/features

- High-precision iRT (non-linear)
- Anchor points

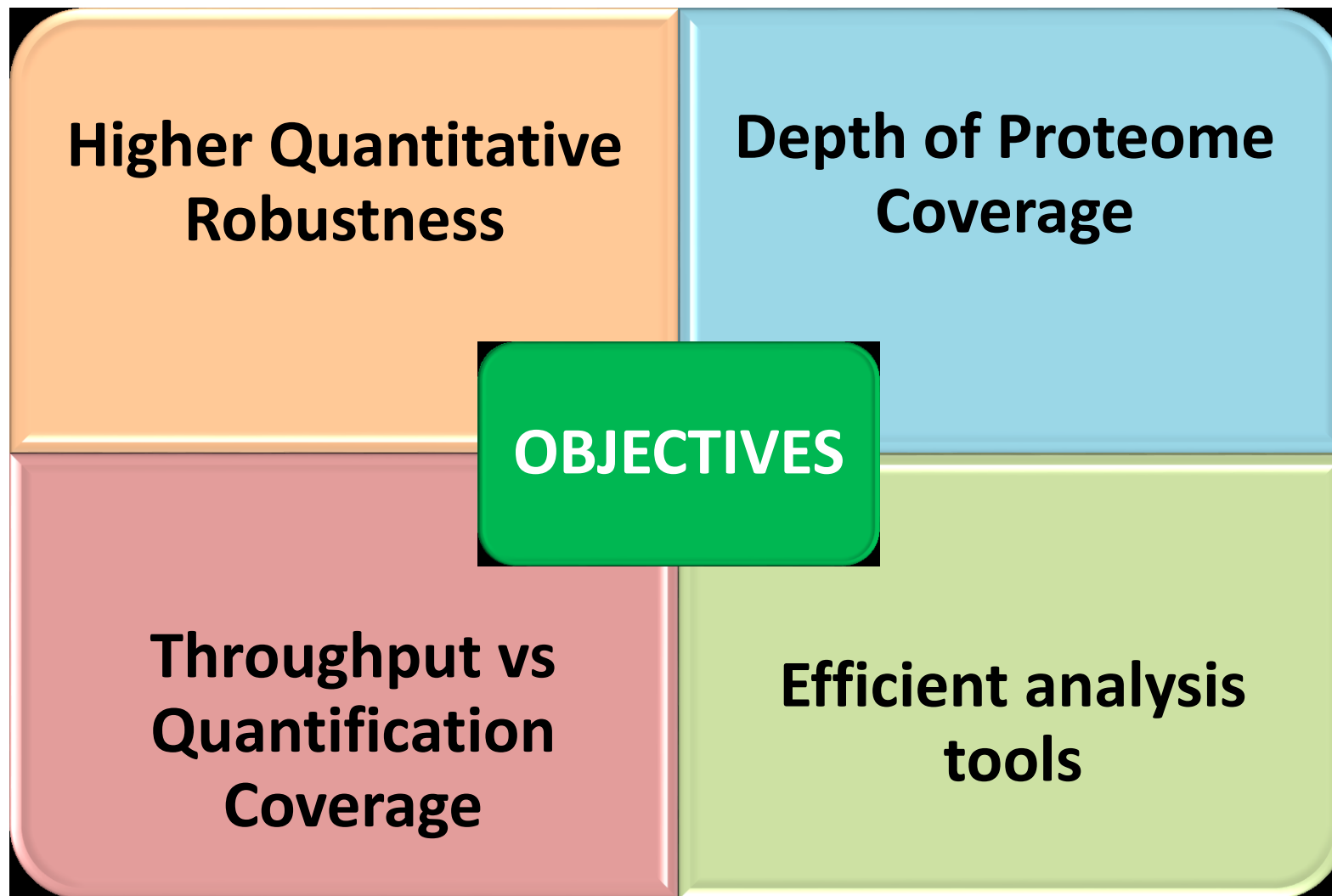


DIA/SWATH-MS: Library based and library free approach



Outline

- DIA/SWATH-MS basic concepts
- **DIA/SWATH-MS objectives and workflows**
- DIA/SWATH-MS analysis tools



Building a Variable or Fixed Window SWATH method in Triple TOF 6600

- Any window strategy can be constructed in text file format and loaded into SWATH Acquisition method editor.
- For Fixed window, manually enter the mass range, SWATH width and accumulation time for each isolation m/z window.

A

Q1 Start	Q1 Stop	CES
399.5	406.5	5
405.5	412.5	5
411.5	418.5	5
417.5	424.5	5
423.5	430.5	5
429.5	436.5	5
435.5	442.5	5

B ↓ Create SWATH Experiments ↑

C

Acquisition Method

- Mass Spectrometer 150.008 mins
 - Period 150.008 mins
 - TOF MS (+)
 - Product Ion (+) 399.5 – 406.5
 - Product Ion (+) 405.5 – 412.5
 - Product Ion (+) 411.5 – 418.5
 - Product Ion (+) 417.5 – 424.5
 - Product Ion (+) 423.5 – 430.5
 - Product Ion (+) 429.5 – 436.5
 - Product Ion (+) 435.5 – 442.5

Create SWATH™ Experiments

Quick Manual

SWATH Analysis Parameters

Start Mass (Da) 400 Stop Mass (Da) 1250 SWATH Width (Da) 10.00 No. of SWATH Per Cycle 85

Fragmentation Conditions

Rolling Collision Energy Collision Energy (V) 10 CES (V) 5

Charge State +1 +2 +3

SWATH Detection Parameters

Start Mass (Da) 100 Stop Mass (Da) 1500 Accumulation Time (ms) 25.00 Total Cycle Time (s) 2.225

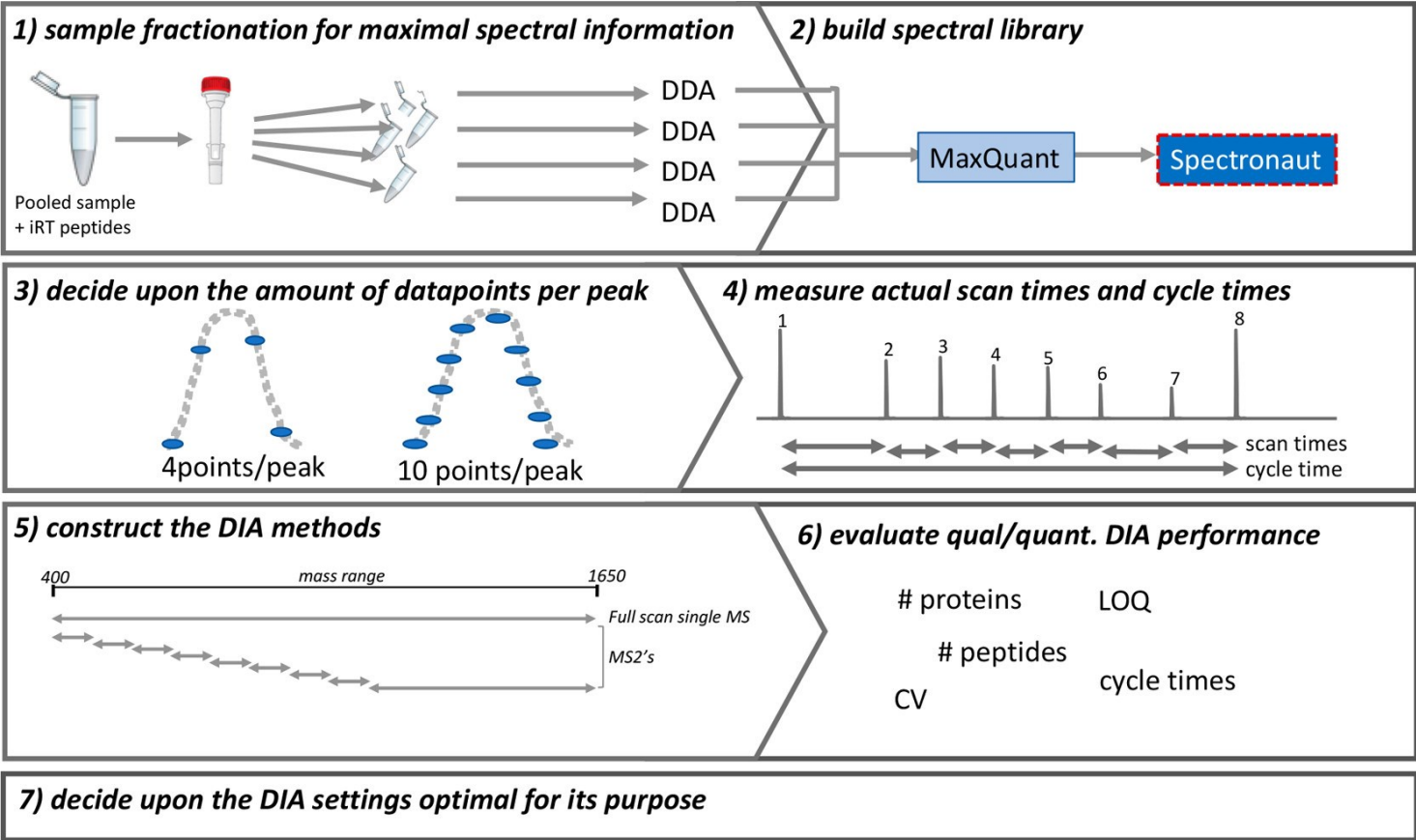
High Resolution High Sensitivity

Read SWATH Windows from Text File

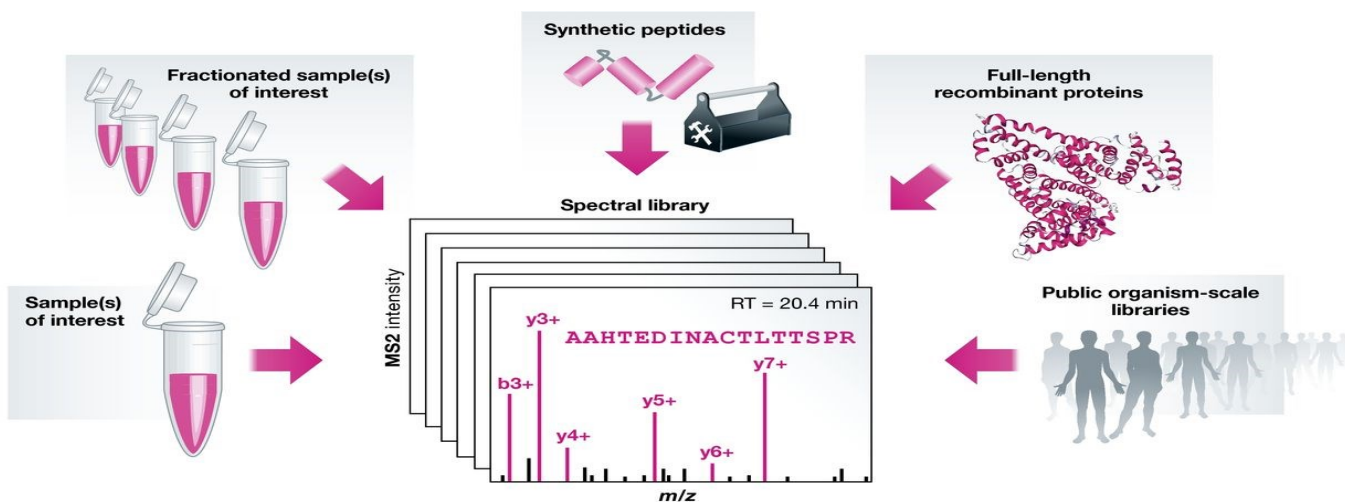
Browse

OK Cancel

DIA workflow for Orbitrap Q Exactive HF



Multiple sources of spectral ion library



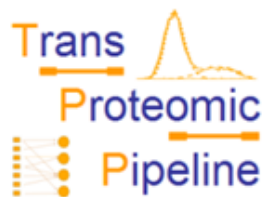
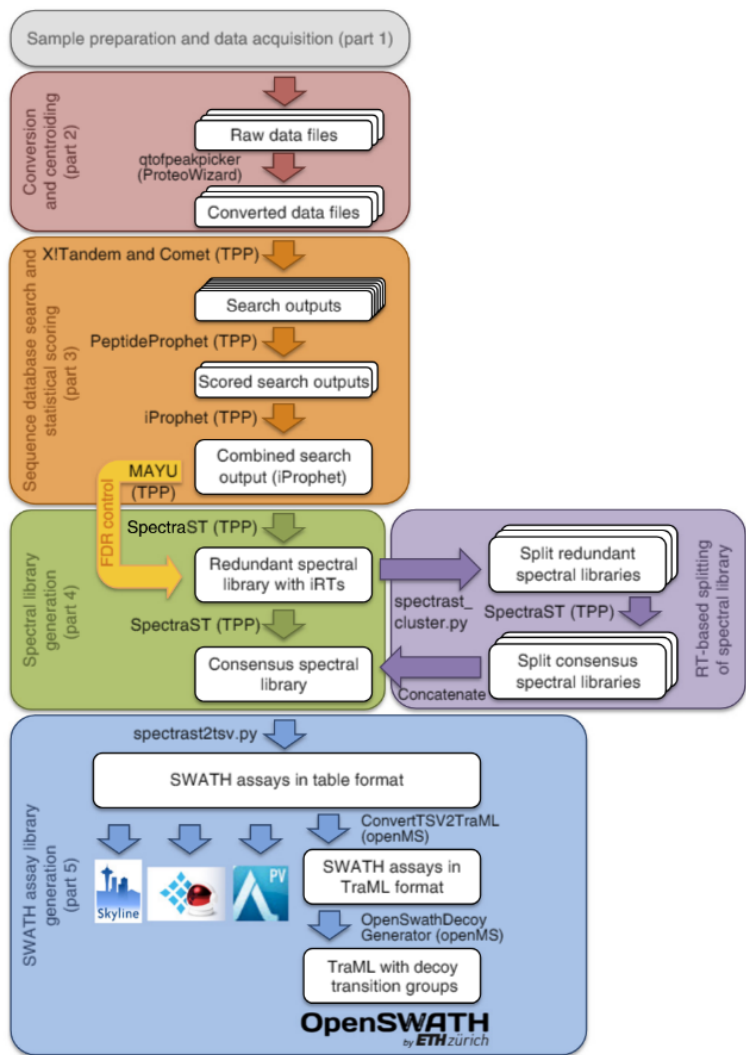
- 4–6 top fragment ions per precursor
- Precursor charge state 2+, 3+, 4+
- Fragment ion charge state 1+, 2+, 3+
- y- and b-ions
- Exclude fragment ions falling into the SWATH precursor isolation window

Peptide query parameters (PQPs)
for target peptide AAHTEDINACTLTTSPR

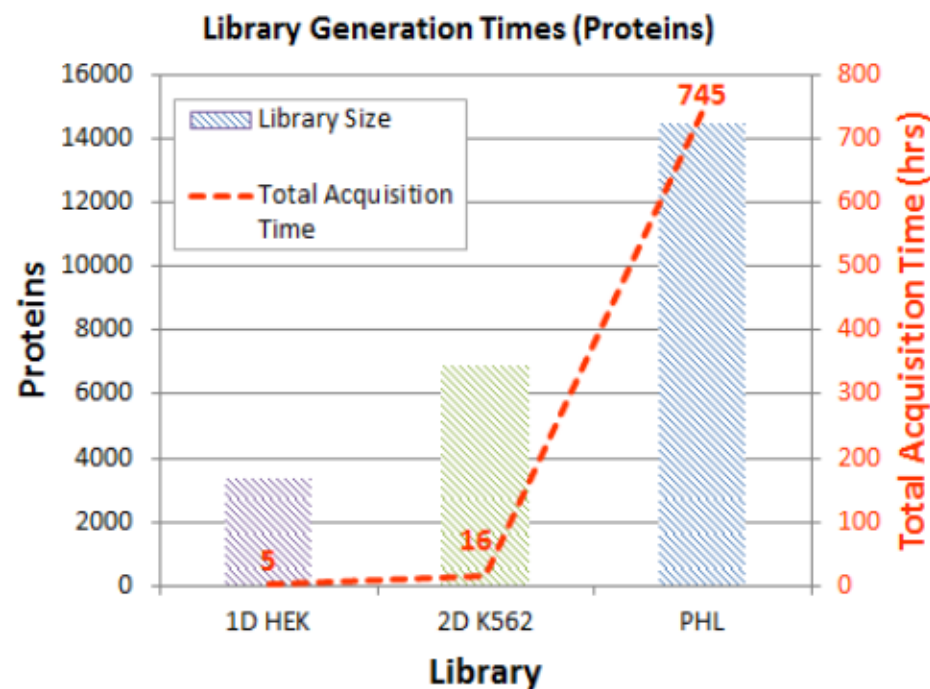
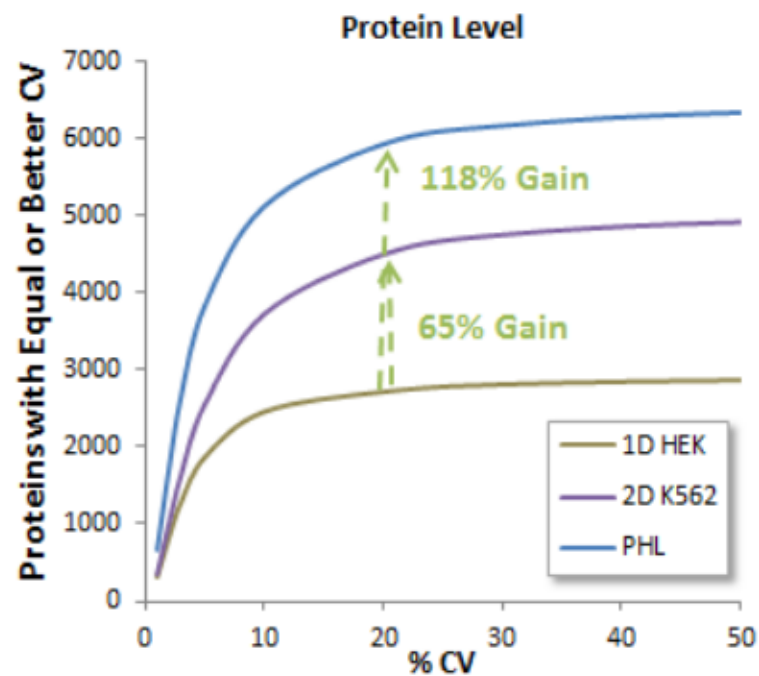
Protein	Peptide sequence	Precursor ion m/z	Fragment ion m/z	Precursor charge	Fragment charge	Fragment ion type	Fragment relative intensity	Normalized retention time
Raf1	AAHTEDINACTLTTSPR	929.4416	775.4308	2	1	y7	70	45.3
Raf1	AAHTEDINACTLTTSPR	929.4416	674.3832	2	1	y6	10	45.3
Raf1	AAHTEDINACTLTTSPR	929.4416	561.2991	2	1	y5	40	45.3
Raf1	AAHTEDINACTLTTSPR	929.4416	460.2514	2	1	y4	20	45.3
Raf1	AAHTEDINACTLTTSPR	929.4416	359.2037	2	1	y3	100	45.3
Raf1	AAHTEDINACTLTTSPR	929.4416	280.1404	2	1	b3	60	45.3

© EMBL

Workflow for DIA/SWATH assay library generation



Impact of Deeper Ion Libraries on Extraction of Quantitative Data from Human Cell Lysate SWATH



SWATH Atlas

Repository for SWATH-MS spectral libraries and results

The screenshot shows the SWATH Atlas website interface. At the top, there is a navigation bar with tabs for Search, All Builds, Current Build, Queries, SRMAtlas, PTPAtlas, Submission, and SWATH/DIA. Below this is a sub-navigation bar with tabs for SWATH Libraries, Custom Libraries, Search datasets, and Virtual MS. The main content area features a table with three columns: Library, Proteome Coverage, and Contributors. The table lists 11 different spectral libraries from various species and their corresponding proteome coverage percentages and protein counts. On the left side of the interface, there are several menu sections: SWATHATLAS HOME, DATA ACCESS (Download Libraries, Customize Libraries), BACKGROUND (Publications, Data Contributors, About SWATH MS), RELATED RESOURCES (Peptide Atlas, SRM Atlas, PASSEL, PASS), and LOGIN. At the bottom left, the logo for the Institute for Systems Biology is visible. In the center, the website URL www.swathatlas.org and the text "11 species" are displayed.

Library	Proteome Coverage	Contributors
H sapiens 10,000 protein 2014	51%, 10,316 of 20218 proteins	Rosenberger et al 2014
H sapiens HLA subtypes		Caron et al 2014
Human breast cancer library 2017		Bouchal et al. 2017
M tuberculosis comprehensive library 2015	99%, 3984 of 4012 proteins	Schubert et. al 2015
Methylobacterium extorquens PA1 - comprehensive library 2016	71%, 3440 of 4829 proteins	Müller et al., 2016
Mouse Immunopeptidome		Caron et al 2018
Pseudomonas syringae DC3000 - comprehensive library 2016	62%, 3406 of 5481 proteins	Müller et al., 2016
Rat (Sertoli cell)	14.6%, 3127 of 21,464 proteins	Dr. Kamal Mandal et. al
S aureus HG001 SWATH Atlas	72% (2091 of 2891 proteins)	Michalik S, Schmidt F, et al
S cerevisiae synthetic peptide full proteome 2013	98%, 6475 of 6607 proteins	Picotti et al 2013
Sphingomonas melonis Fr1 - comprehensive library 2016	71%, 2729 of 3857 proteins	Müller et al., 2016
Zebrafish multi-tissue 2019	40.4%, 10,405 of ~25,500 proteins (ensembl)	Blattmann et al

www.swathatlas.org
11 species

DIA/SWATH- Prior Knowledge and Consistency

- Relative product ion abundance
 - Spectral ion libraries
- Retention time Information
 - iRT libraries
- Powerful enough to be used cross-lab/ cross experiment.
- More powerful run to run compared to DDA.

Outline

- DIA/SWATH-MS basic concepts
- DIA/SWATH-MS objectives and workflows
- **DIA/SWATH-MS analysis tools**

DIA/SWATH-MS Data Analysis tools

Commercial

<p>Sciex</p>	<p>Biognosys</p>	<p>Proteome Software</p>	<p>Bioinformatics Solutions</p>

OPEN SOURCE

<p>Institute for Systems Biology</p>	<p>Data Independent Signal Correlator ISB</p>	<p>ETH Zurich</p>	<p>U. Michigan</p>	<p>U. Washington</p>

Peak Scoring and Picking Models

- mProphet algorithm is implemented in OpenSWATH, Skyline and Spectronaut.
- PeakView scoring is based on spectral (correct isotope, m/z error, MS2 quality) and chromatographic attributes (RT, Peak width).
- Available feature scores are different that contribute to discriminant (d-score) or Composite (c) score.

Name: test

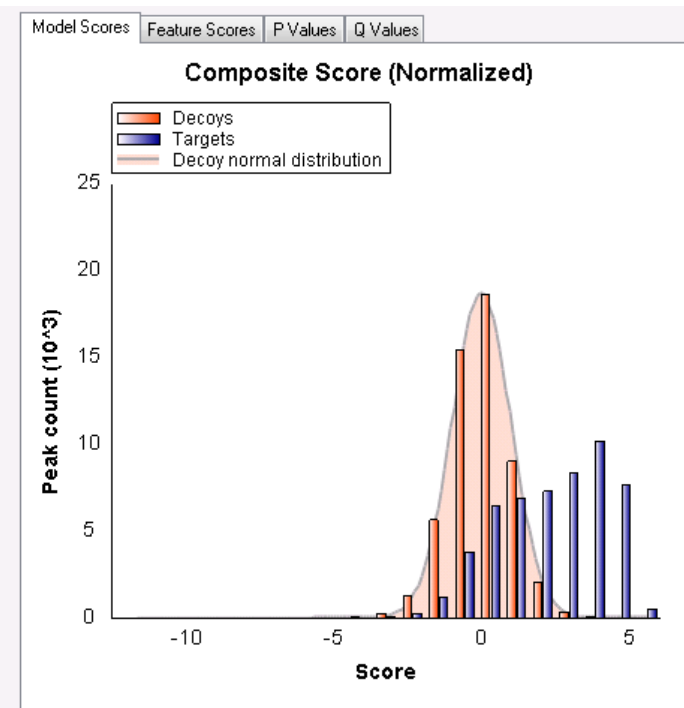
Choose model: mProphet

Training

Use decoys Use second best peaks

Available feature scores:

Enabled	Score Name	Weight	Percentage Contribution
<input checked="" type="checkbox"/>	Intensity	-0.3726	-5.5%
<input checked="" type="checkbox"/>	Retention time difference	-0.2938	6.7%
<input checked="" type="checkbox"/>	Retention time difference squared	-0.0043	0.8%
<input checked="" type="checkbox"/>	Library intensity dot-product	5.7457	28.6%
<input checked="" type="checkbox"/>	Shape (weighted)	3.4476	22.8%
<input checked="" type="checkbox"/>	Co-elution (weighted)	-0.2301	16.2%
<input checked="" type="checkbox"/>	Co-elution count	0.4843	12.2%
<input checked="" type="checkbox"/>	Signal to noise	0.4022	5.3%
<input checked="" type="checkbox"/>	Product mass error	-0.2300	12.8%



Acknowledgements

Proteomics Research Group (PRG)

Mission



The Proteomics Research Group (PRG) is a volunteer scientific organization dedicated to sharing knowledge about the analysis of proteins. The PRG aims to assist protein scientists and resource facilities

to achieve their highest potential by sponsoring annual research studies that examine current techniques and capabilities. Through the promotion of broad participation and scientific excellence, the PRG aims to raise awareness, knowledge and education about modern methods of protein analysis.

Current Membership

Pratik Jagtap (Chair)

University of Minnesota

Laura Herring

University of North Carolina at Chapel Hill

Joanna Kirkpatrick

Leibniz Institute on Aging – Fritz Lipmann Institute

Mukul Midha

Institute of Systems Biology

LeRoy Martin

Waters Corporation

Benjamin Neely

National Institute of Standards and Technology

Brett Phinney

University of California Davis

Baozhen (Paul) Shan

Bioinformatics Solutions Inc.

Paul M. Stemmer

Wayne State University

Yan Wang

University of Maryland

Allis Chien (EB Liaison)

Stanford University

<https://abrf.org/research-group/proteomics-research-group-prg>