«Développement de méthodes de quantification ciblée de peptides (SRM, PRM et DIA) et applications »

23 Mars 2017

Christine CARAPITO
ccarapito@unistra.fr

Laboratoire de Spectrométrie de Masse Bio-Organique
LSMBO – IPHC – UMR7178

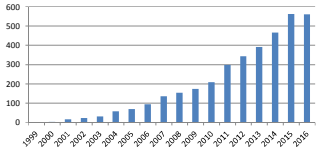
1

Targeted proteomics

Targeted Proteomics approaches gain

increasing importance in the overall proteomics toolbox !

Pubmed : « Targeted proteomics » publications



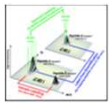
- High sensitivity, selectivity and quantitative accuracy
- Large dynamic range
- High reproducibility
- Reasonable (increasing) multiplexing capability

2

From shotgun to targeted proteomics

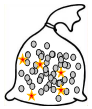
Discovery, Shotgun Proteomics

- ✓ Extensive fractionation methods: Depletion, Enrichment, 1D-2D Gels, Multiple chromatographies, …
- ✓ Shotgun, DDA/DIA LC-MSMS approaches
- ✓ Relative global quantitation
  - Label-free
  - Isotopic labeling
  - Spectral counting



From Mueller, L. N., et al., 2008

LC-MS/MS

Qualitative	Quantitative
 500-8000 identified proteins	 Approx. relative quantitation

Targeted Proteomics

Targeted MS

Heavy labeled synthetic standards

LC-SRM
LC-PRM

Qualitative	Quantitative
 10-100 candidate proteins	 Reproducible, precise (absolute) quantitation

3

From shotgun to targeted proteomics

Targeted MS methods are fairly robust and powerful

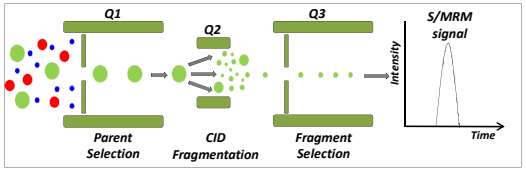
But still many struggle with experimental design and data analysis

4

The historical targeted approach :

LC-**S**elected/**M**ultiple **R**eaction **M**onitoring (S/MRM)

on **triple quadrupole**-type instruments (QqQ, Q-Trap)



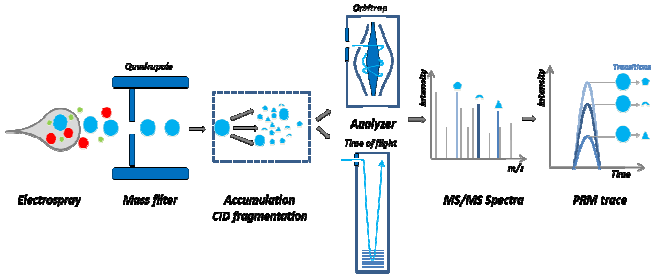
Come back of QqQ instruments in proteomics labs in 2005-2006

- *Lange V et al., Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4:222.*
- *Picotti P. et al., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods. 2012;9(6):555-66.*

5

HRAM Targeted approaches

Since 2012, use of high-resolution accurate mass analysers for targeted acquisition



Transitions are extracted post-acquisition

« Sophisticated inclusion lists »

Bourmaud A., *et al,* Proteomics (2016) ; Schilling B., *et al,* Anal Chem (2015)

6

HRAM Targeted approaches

Parallel Reaction Monitoring on Q-Orbitrap instruments

- Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., and Coon, J.J. (**2012**)
Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics, *Mol Cell Proteomics 11(11), 1475-88.*
- Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (**2012**)
Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer, *Mol Cell Proteomics 11(12), 1709-23.*

HR MRM on Q-TOF instruments from ABSciex (TripleTOF)

-Tong, L., Zhou, X.Y., Jylha, A., Aapola, U., Liu, D.N., Koh, S.K., Tian, D., Quah, J., Uusitalo, H., Beuerman, R.W., and Zhou, L. (**2015**)
Quantitation of 47 human tear proteins using high resolution multiple reaction monitoring (HR-MRM) based-mass spectrometry, *J Proteomics 115, 36-48.*

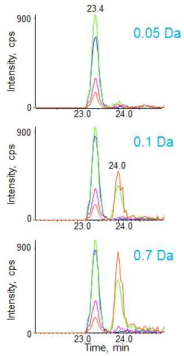
TOF-MRM or HD-MRM (HD for High Definition, with mobility separation) on Waters Q-TOF

High resolution MRM on Bruker Q-TOF

7

PRM versus SRM

- Easier method development (DDA data acquired on the same instrument)
- Full MS2 spectra: All fragments (transitions) measured: specificity
- Removing Interferences or Background through High Resolution Fragment Ion Extractions.



8

Data-Independent Acquisition (DIA)

Historically,

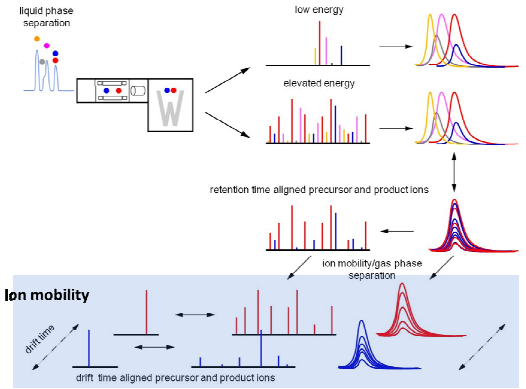
- Venable et al. Nat. Methods 1, 39–45 (2004), Thermo LTQ linear Ion Trap

- Waters MS^E strategy
  Moran, D., et al., J Virol Methods, 2014. 195:9-17.

9

WATERS MS^E strategy

Waters has a complete workflow since 2006, including identification algorithms



PLGS

10

DIA SWATH

Data Independent Acquisition strategy really gained increased interest with the introduction of SWATH-Acquisition on Q-TOF (ABSciex TripleTOF) instruments in 2012:

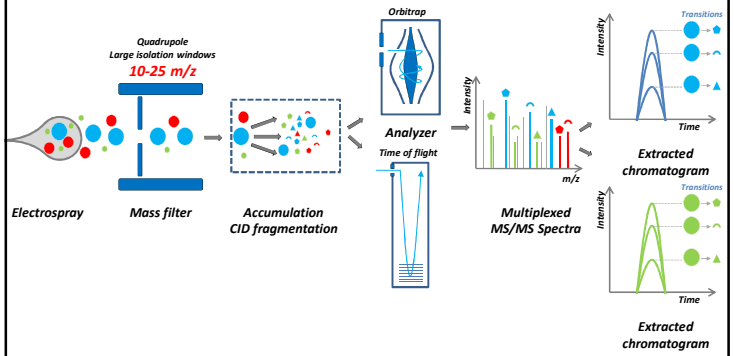
Gillet et al., (2012), Mol Cell Proteomics 11(6), O111 016717

Based on a commercial instrument but developed in academia, with open-source dedicated software development !

https://www.youtube.com/watch?v=Ixa4m-0BsRU

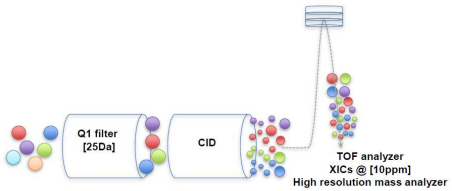
11

Data-Independent Acquisition



DIA SWATH

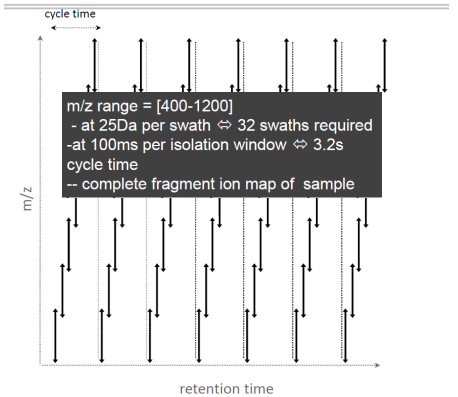
SWATH-MS: Sequential Windowed data independent Acquisition of the Total High-resolution Mass Spectra



13

DIA SWATH

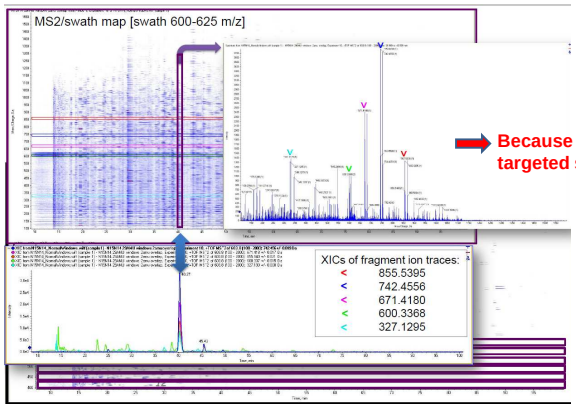
SWATH-MS Acquisition Principle



14

DIA SWATH

Why is SWATH a targeted method today?

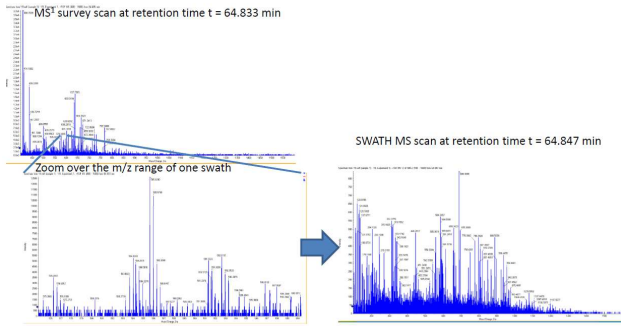


Because we extract targeted signals

15

DIA SWATH

SWATH-MS data is incompatible with conventional database searching

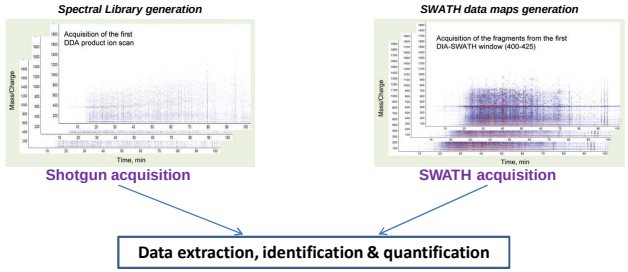


Composite spectra of high complexity

16

DIA SWATH

Targeted signal extraction, based on a spectral library

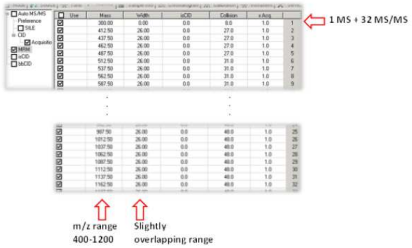


DIA-Umpire strategy : non targeted DIA data interpretation

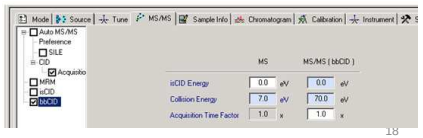
Tsou, C.-C. et al. (2015) Nat. Methods ,12, 258–264.

DIA on Bruker Instruments

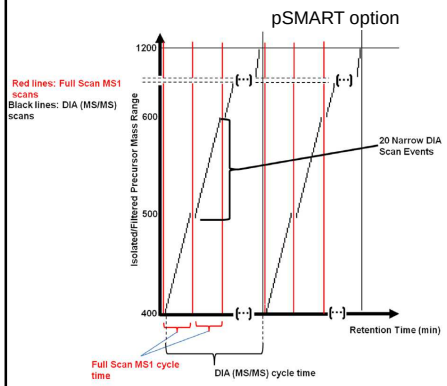
- Middle-band CID: equivalent to SWATH acquisition (Setup in HR MRM mode)



- Broad-band DIA: Equivalent to MSE



DIA on Thermo Instruments



On Q-Exactive Instruments

- MS1 library quantification
- Confirmation of Identity using narrow fragmentation windows and library
- User defined time between 2 MS1 scans: direct impact on quantification accuracy

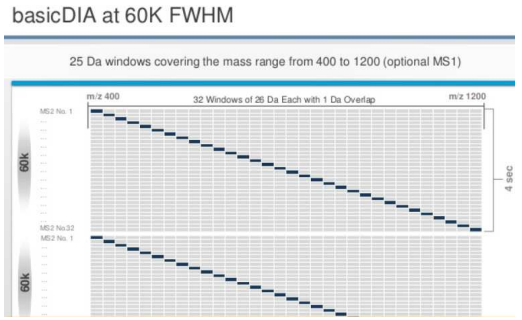
Prakash et al. 2014, J Prot Research, 13 (12), 5415–5430

DIA on Thermo Instruments

*On Q-Exactive **HF***

SWATH-type DIA

DIA
- 26 amu isolation
- Variable first mass
- 60 000 resolution
- AGC: 2e5
- 105ms Injection
- NCE 28%
- Loop Count 32

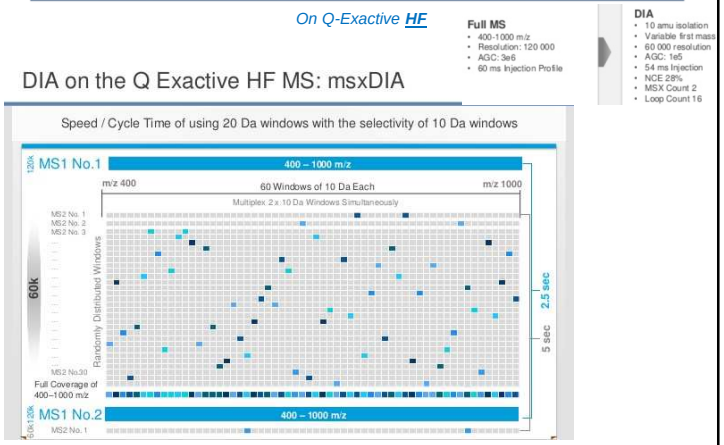


Egertson et al. 2013, Nature Methods, 10, 744-766

DIA on Thermo Instruments

*On Q-Exactive **HF***

DIA on the Q Exactive HF MS: msxDIA

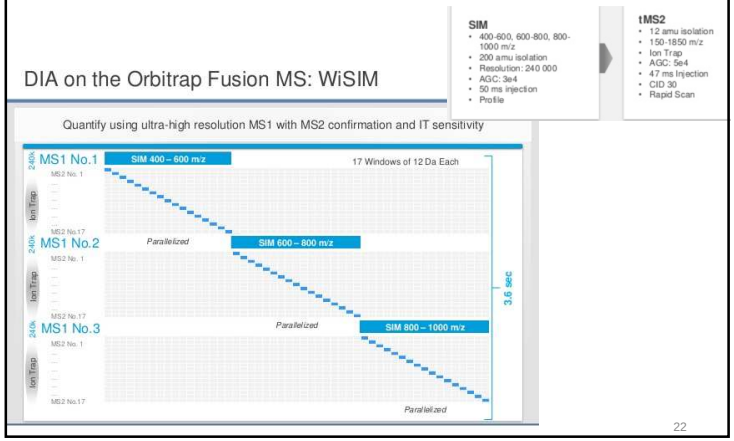


Egertson et al. 2013, Nature Methods, 10, 744-766

DIA on Thermo Instruments

On Fusion or Lumos

DIA on the Orbitrap Fusion MS: WiSIM



Data Independent Acquisition (DIA)

The major limitation resides in

data interpretation tools and

robust data processing workflows

are missing !

Important considerations for targeted assays

1. Multiplexing capacity

   ➡ 3, 50 proteins, or largest possible spectral library?

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity

   µg/ml level in unfractionated plasma
   Anderson, L. et al., Mol. Cell Proteomics, 2006; Kuzyk, M.A. et al. Mol. Cell Proteomics, 2009; Addona, T.A. et al., Nat. Biotechnol., 2009; Dutta, A., et al., Oncotarget, 2016

   ng/ml level in fractionated plasma (immunoaffinity depletion, SISCAPA, …)
   Keshishian, H., et al., Mol. Cell Proteomics, 2007; Fortin, T. et al. Mol. Cell Proteomics, 2009; Anderson, N.L. et al. J. Proteome Res., 2004; Karakosta, T.D., et al., Mol Cell Proteomics, 2016

   amol level routinely detected in complex samples
   Shi T., et al., Proteomics, 2012; Schiess, R., et al., Mol. Oncol., 2012; Frottin, F., et al., Oncotarget, 2016, …

25

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'

   Stable-isotope dilution

   ➡ isotopically labeled reference peptides : AQUA (Absolute Quantification)
   (Gerber, S.A. et al., Proc. Natl. Acad. Sci. USA, 2003)

   ➡ concatenated tryptic peptides : QconCATs (Quantification concatamer)
   (Beynon, R.J., et al., Nat. Methods , 2005)

   ➡ full length isotope-labeled proteins: PSAQ (Protein Standard Absolute Quantification)
   (Dupuis, A., et al., Proteomics , 2008)

26

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'
4. Reproducibility

   ➡ Sample preparation (limited fractionation)
   ➡ Chromatography (scheduling, retention time correction)
   ➡ MS analysis
   CVs < 15-20%

27

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'
4. Reproducibility
5. Sample throughput

   ➡ Up to 100 proteins/hour
   Up to 24 runs/day

28

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'
4. Reproducibility
5. Sample throughput
6. Robustness
   - ➡ Frequence of instrument cleaning
   - ➡ Chromatography scale (nano-, micro-, normal-flow rates)

29

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'
4. Reproducibility
5. Sample throughput
6. Robustness
7. Quantification of PTMs, discrimination of isoforms
   - ➡ Phosphorylation
     Chan, C.Y., et al., Expert Rev Proteomics, 2016
   - ➡ Acetylation, propionylation, methylation and ubiquitination
     Darwanto, A. et al., J. Biol. Chem., 2010 ; Philp, A., et al., Am J Physiol Cell Physiol., 2014
   - ➡ Isoform discrimination: Multiple peptides = multiple "antigens" per protein
     Wang, Q. et al., Proc. Natl. Acad. Sci. USA, 2011; Boja, E.S., et al., Clin Proteomics, 2014

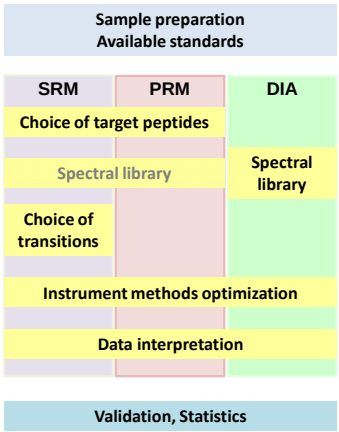
30

Important considerations for targeted assays

1. Multiplexing capacity
2. Sensitivity
3. Absolute quantification, 'precise relative quantification'
4. Reproducibility
5. Sample throughput
6. Robustness
7. Quantification of PTMs, discrimination of isoforms
8. Method transferability (intra- inter-laboratory)

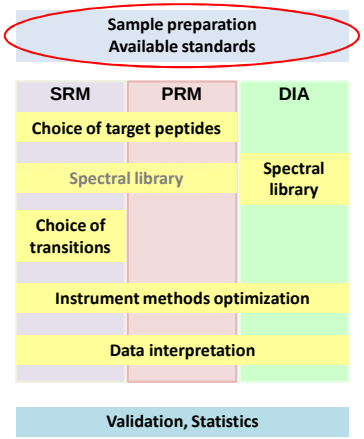
31

Assay development workflow



32

Assay development workflow



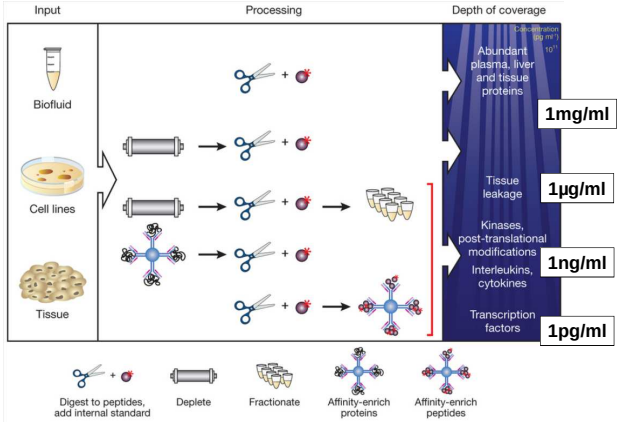
33

Sample Preparation

- Each sample has its own story
- Same recommandations as for global proteomics
- But more crucial when quantification is required

34

Sample prefractionation



From Gillette and Carr, Nature Methods, 2013

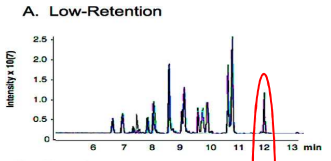
35

Precautions

- **Reproductibility in sample preparation** crucial
- **Biological replicates** : min 3 but 5 is better
- **Limit sample preparation, prefractionation steps**
- **Ideally fresh samples**, absolutely avoid freeze/thawing cycles
- **Avoid precipations** (solubilisation problems in buffers compatible with further proteomics prep (trypsin, MS, …)
- **Careful with containers**
- **Stability study, only for clinical studies**
  - ➔Hardly applicable in fundamental research programs
    - Not enough replicates
    - Low amounts of material

36

Sample storage/containers



Differential Recovery of Peptides from Sample Tubes and the Reproducibility of Quantitative Proteomic Data Steven J. Bark Journal of Proteome Research 2007, 6, 4511-4516

Peptide storage: are you getting the best return on your investment? Defining optimal storage conditions for proteomics samples. Kraut A. et al. *J Proteome Res.* (2009) 8(7):3778-85.

37

Available Standards

1/ Internal heavy labelled standards for isotopic dilution

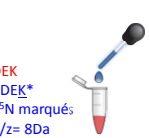
- Crude labeled peptides: very useful for method development and precise relative quantification
- AQUA peptides: necessary when « absolute » quantification is needed (generally in a second step)
- QconQat
- PSAQ proteins : ideal to evaluate also protein digestion efficiency, high cost

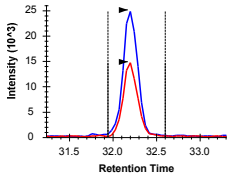
38

Heavy labeled peptides/proteins

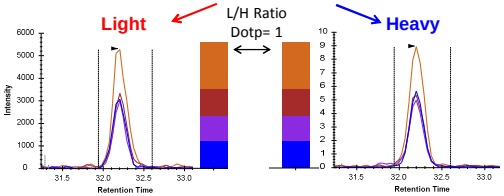
Isotopic dilution: Use of heavy labeled synthetic standards
AQUA, QconCATs, PSAQ

Light: PEPTIDEK
Heavy: PEPTIDEK*
^{13}C et ^{15}N marqués
Δm/z= 8Da





- Relative intensities

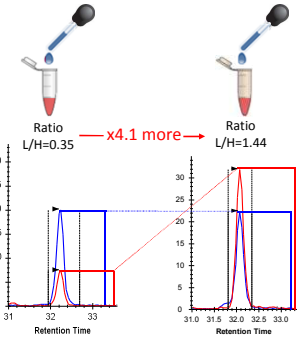


39

Heavy labeled peptides/proteins

Isotopic dilution: Use of heavy labeled synthetic standards
AQUA, QconCATs, PSAQ

Light: PEPTIDEK
Heavy: PEPTIDEK*
^{13}C et ^{15}N marqués
Δm/z= 8Da



- Precision
- Accuracy
- Relative quantification, even Absolute

40

Heavy labeled peptides/proteins

Different vendors, different grades, purities, prices,… qualities!

PEPotec SRM Peptide Libraries
Three grades to fit your experimental needs

	Grade 1 Fast and Easy	Grade 2 Greater Analysis	Grade 3 Maximum Assurance
Quantity	>0.1mg		
Length[1]	6 to 25 amino acids		
Purity	Crude (as synthesized)		
Formulation[1]	Suspended in 0.1% TFA in 50% (v/v) acetonitrile/water		
Delivery format	Thermo Scientific Matrix 96-tube plates (Product # 3712MTX)		
C-terminal residue[1]	R or K		
Counter ion	TFA		
Quality control (QC)	MS check of 5% of peptides	MS check of 100% of peptides	MS analysis of 100% of peptides
Peptide resynthesis[2]	Not provided	Not provided	One resynthesis provided
Failed synthesis policy	You pay for entire set of peptides ordered	You pay only for peptides synthesized	You pay only for peptides synthesized
Included documentation	Peptide amount	Peptide amount	Peptide amount and MS spectra
Minimum order[3]	24 peptides	4 peptides	4 peptides

1 Changes to the standard length restrictions, formulation and C-terminal residues are available as optional services
2 Peptides not detected during MS analysis will be resynthesized (depending on the grade selected)
3 Orders for fewer than 24 peptides incur a plate fee

Thermo

PEPscreen Service Specifications

Sigma

Library Size	24 peptide minimum
Peptide Length	6 to 20 amino acids
Quantity	0.5 - 2mg or 2 - 5mg
Peptide Form	Dried film at the bottom of individual tubes
N-Terminal	Free amine or acetylated
C-Termini	Free acid or amidated
Non-standard Amino Acids	Any commercially available non-standard amino acid
Random Library	Any mixture of commercially available amino acids
Chemical Modifications	Cyclization, phosphorylation, biotinylation, PEGylation, acylation, etc.
Dye Labeling	Flc, FITC, Dansyl, Dabcyl, Dabsyl, TAMRA™, Lissamine, etc.

SpikeTides™_L - isotopically labeled peptides

SpikeTides™_L Proteotypic Peptides

SpikeTides™_L are isotopically labeled, proteotypic peptides that terminate with C-terminal heavy Arg/Lys.

Amounts: appr. 10-30nmol/peptide (exact amount unknown!)
Purity: Unpurified
Application: Development of SRM assays and relative quantification of proteins using a single product
Price: from 19 US$ / 17 € (depending on number of peptides ordered)
Validated pooling service available upon request!

JPT

SpikeTides™_L

The use of this product may be subject of U.S. patent No. 7,501,286, the European patent 1,472,539 and corresponding patent applications owned by the President and Fellows of Harvard College, USA. The purchase of this product does not convey a license under any method claims in the foregoing patents or patent applications.

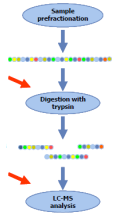
Available Standards

Additional useful standards for quality control, method development

To minimize the analytical variability, a number of quality control (QC) products have been developed to evaluate the efficacy of individual steps within a bottom-up proteomic experiment

Collective processes DIGESTIF workflow QC kit
QCAL standard for MS assessment
Reversed phase liquid chromatography calibrant (RePLiCal)

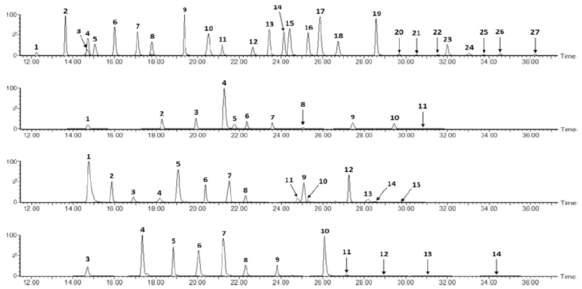
Indexed Retention Time (iRT) kit for RT identification and correction
Reversed phase liquid chromatography calibrant (RePLiCal)
Halogenated peptides as internal standards (H-PINS)



42

Standards for Quality Control

Mixtures of standard peptides to spike in all samples



RePLiCal, PolyQuant

iRT, Biognosys

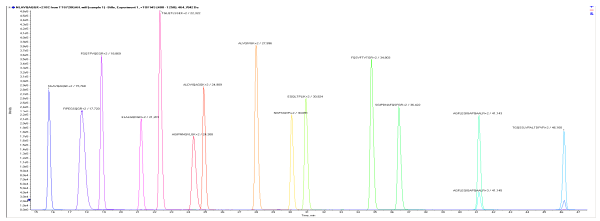
Pierce retention time calibration mixture (Thermo Scientific)

MS RT calibration mix (Sigma-Aldrich)

43

Standards for Quality Control

+ Kit READYBEADS (Anaquant)

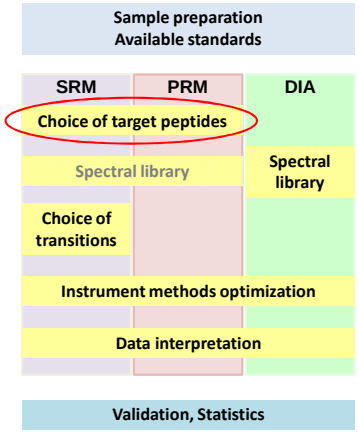


Very useful for :

- Method transferability across gradients and instrumental platforms
- Testing of trapping column performances
- Detection in very complex matrices
- Prediction of RTs
- Optimize LC conditions, …

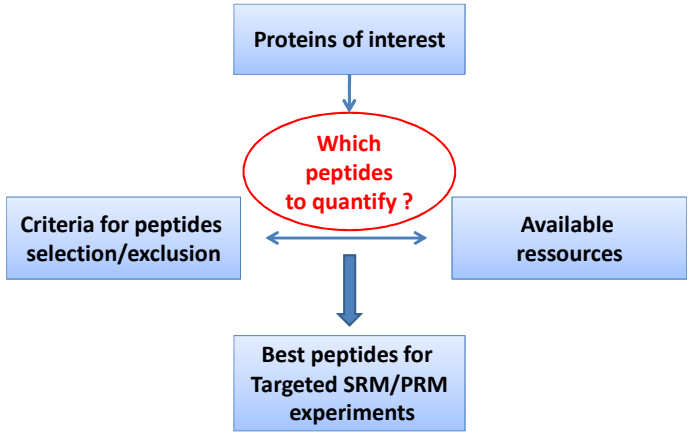
44

Assay development workflow



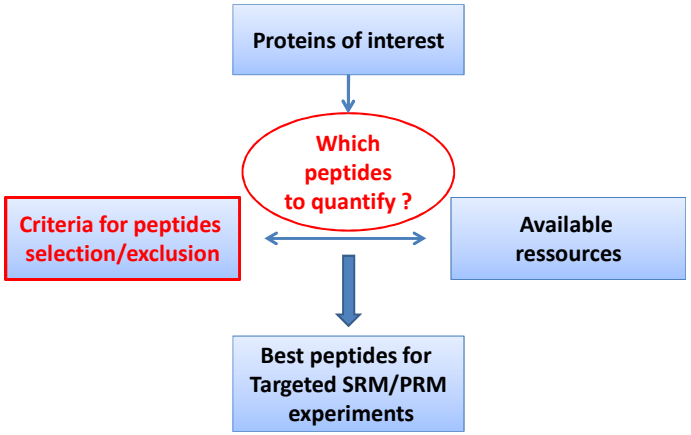
45

Workflow



46

Workflow



47

Criteria for peptides selection / exclusion

Proteotypicity

- Unicity for the protein it is the signature of
- Visible in MS

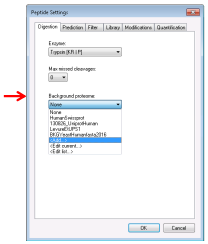
1) The same peptide sequence can be shared by several proteins
Skyline : add a background proteome

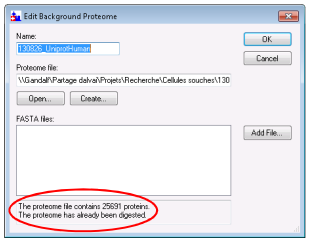
48

Criteria for peptides selection / exclusion

Skyline: background proteome

- Fasta file containing all proteins of your sample
  (i.e. : all human sequences from swissprot)
- Skyline : Peptide settings/Digestion/Background proteome
  From your Fasta file, Skyline creates a .blib file





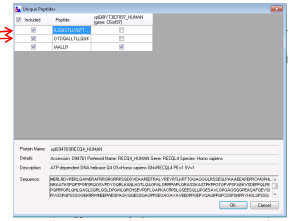
Skyline

Criteria for peptides selection / exclusion

Skyline: background proteome

- Fasta file containing all proteins of your sample
  (i.e. : all human sequences from swissprot)
- Skyline : Peptide settings/Digestion/Background proteome
  From your Fasta file, Skyline creates a .blib file
- You can check the proteotypicity of all peptides of your proteins

Proteotypic



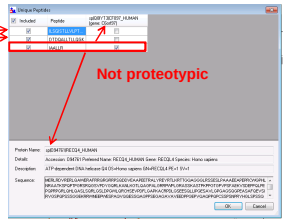
Skyline

Criteria for peptides selection / exclusion

Skyline: background proteome

- Fasta file containing all proteins of your sample
  (i.e. : all human sequences from swissprot)
- Skyline : Peptide settings/Digestion/Background proteome
  From your Fasta file, Skyline creates a .blib file
- You can check the proteotypicity of all peptides of your proteins

Proteotypic

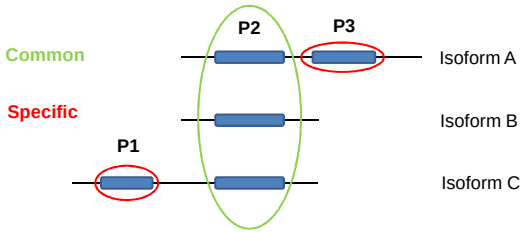


Skyline

Criteria for peptides selection / exclusion

Proteotypicity

2) When a protein has several isoforms, which one is quantified ?



Criteria for peptides selection / exclusion

Proteotypicity

2) When a protein has several isoforms, which one is quantified ?

3) Canonic sequence or not?

The given sequence in Uniprot corresponds to the most common polymorphic variant but others can exist.

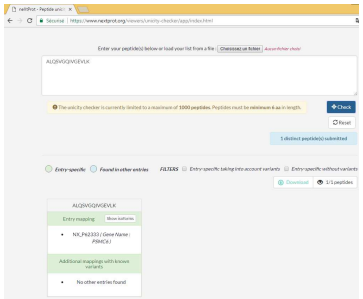
> 3 million sequence variants annotated in SwissProt

Criteria for peptides selection / exclusion

Proteotypicity

2) When a protein has several isoforms, which one is quantified ?

3) Canonic sequence or not?



Unicity checker of neXtProt
(integrates sequence variants, isoforms and I/L null switch):
https://www.nextprot.org/viewers/unicity-checker

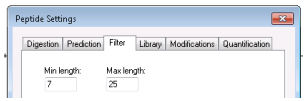
Criteria for peptides selection / exclusion

PEPTIDE SIZE

Ideally peptides should contain between 7 and 25 amino-acids

Short peptides (< 7 aa):
- Less specific
- Less fragments for transitions choice, more interferences

Long peptides (>25 aa)
- Instrument mass range limits (6500Qtrap : max m/z 1250)
- Hydrophobicity
- Synthesis difficulties and price



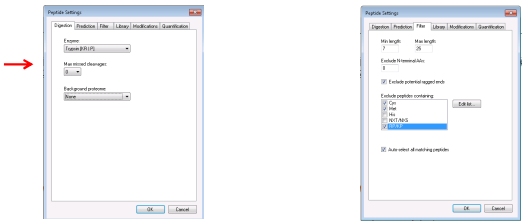
Criteria for peptides selection / exclusion

TRYPSIN CLIVAGE SITES

- Avoid peptide containing **misscleavage** site(s)

- Potential **ragged ends**
  Avoid peptides with KR, RK, RR ou KK before or after the sequence
  (even if you are able to see the fully cleaved peptide)

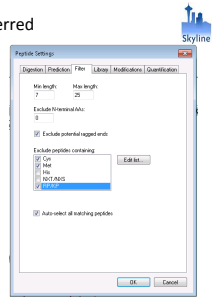
- Avoid peptides containg **KP or RP** : usually no clivage but not always…



Criteria for peptides selection / exclusion

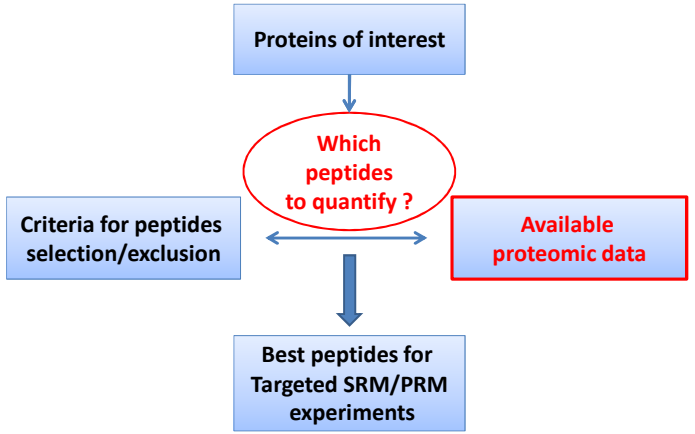
PEPTIDE MODIFICATIONS

- Uncontrolled : **Oxidations of methionines** (and tryptophanes)
  Avoid peptides containing M (and/or W)
- Controlled : **Alkylation of cysteines**
  Can be kept but peptides without cysteines are preferred
- **Post-Translational Modifications**
  (Phosphorylations, glycosylations, …)
  Sites can be described or not (see Uniprot)
  Might be needed



57

Workflow



58

Available proteomic data

Some peptides have better ionization / fragmentation than others.
To select the best peptides for targeted quantification of a protein, use preferentially:

1) Experimental proteomic data already acquired on your samples of interest

59

Available proteomic data

1) Experimental proteomic data on your samples of interest

- **DDA acquisitions (LC-MSMS) on the same instrument**

- Possible with a QTrap instrument which can work in SRM or DDA mode
- Possible with high resolution instruments (Q-TOF or Q-Exactive) working in DDA and then PRM mode

- **DDA acquisitions on other instruments**

- Linear traps, Orbitraps, Q-TOFs….
- More identifications = more peptides in the library

!!! Trap fragmentation is slightly different from quadrupole fragmentation used in SRM
The HCD fragmentation is a quadrupole-like fragmentation
Using the same instrument for DDA and PRM experiments is also ideal in this regard

60

Available proteomic data

Some peptides have better ionization / fragmentation than others.
To select the best peptides for targeted quantification of a protein, use preferentially:

1) Experimental proteomic data already acquired on your samples of interest

2) Experimental proteomic data on other samples in which your protein(s) of interest has/have been identified

61

Available proteomic data

2) Experimental proteomic data on other samples

Any proteomic data in which your proteins of interest have been identified can be useful

Even if samples are very different, what you need is high quality spectra
(= good identification + nice fragmentation)

Make a screening of your own data !!!

62

Available proteomic data

Some peptides have better ionization / fragmentation than others.
To select the best peptides for targeted quantification of a protein, use preferentially:

1) Experimental proteomic data already acquired on your samples of interest

2) Experimental proteomic data on other samples in which your protein(s) of interest has/have been identified

3) Peptides repositories available online

63

Available proteomic data

3) Peptides repositories available online

When no experimental data are available in your lab, use peptide repositories
They are also **created from experimental data**

- ProteomeExchange – Pride *(Vizcaino J.A., et al., Nat Biotechnol, 2014)*
  Proteomic raw data repository
- Peptide Atlas *(Desiere S. et al., Nucleic Acids Research, 2006)*
  Proteomic raw data repository
- SRM Atlas *(Picotti P. et al., Nat. Methods, 2010)*
  Compendium of SRM measuremnts on natural and synthetic peptides
  99% of human proteome (20200 proteins), 80% of mouse proteome (12000 proteins), yeast proteome
- Passel, The PeptideAtlas SRMexperiment library *(Farrah T. et al., Proteomics 2012)*
  SRM raw data repository (SRM experimental results from analysis of biological samples)
- MRMaid Pride repository *(Mead, J.A. et al., Mol.Cell Proteomics, 2008)*
  Peptides suggestions for SRM experiments based on Pride proteomic database
- The GPM : the Global Proteome Machine – databases

64

Available proteomic data

SRM Atlas

Objective:
Prepare targeted methods for all human proteins and optimize SRM assays for 5 proteotypic peptides per human protein

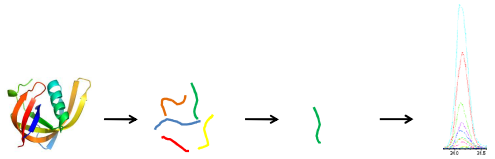
Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome.
Cell. 2016 Jul 28;166(3):766-78.

65

Available proteomic data

SRM Atlas

Uniprot SwissProt contains approx. 20,100 human proteins



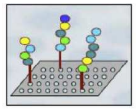
1 protein	5 peptides	1-2 Charge states	3-10 transitions
SRM Atlas			
~20,000	~100,000	~150,000	450,000-1,500,000

Institute of Molecular Systems Biology, IMSB, ETH Zurich (R. Aebersold)

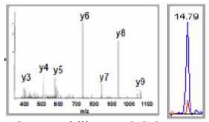
Available proteomic data

SRM Atlas

Choice of 5 proteotypic peptides for each protein



Peptide synthesis
(time & cost)



Spectral library QQQ

Compilation in SRM atlas

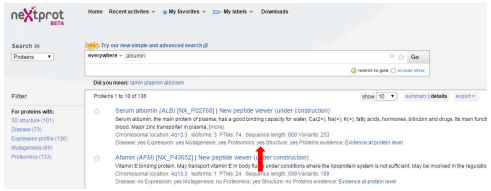
Validated SRM assays, best transitions, retention times, …

ID	Q1	Q3	RT
Pep1	563.28	875.12	16
Pep1	563.28	789.21	16
Pep1	563.28	872.31	16

Available proteomic data

3) Peptides repositories available online - Useful proteomics tools

Nextprot (http://www.nextprot.org/) : PeptideAtlas and SRMAtlas links

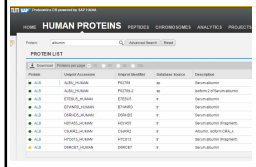


68

Available proteomic data

3) Peptides repositories available online - Useful proteomics tools

Proteomics DB - Human Proteome (https://www.proteomicsdb.org/)



69

Available proteomic data

Some peptides have better ionization / fragmentation than others.
To select the best peptides for targeted quantification of a protein, use preferentially:

1) Experimental proteomic data already acquired on your samples of interest

2) Experimental proteomic data on other samples in which your protein(s) of interest has/have been identified

3) Peptides repositories available online

4) *In silico* prediction of best flyers peptides

70

Available proteomic data

4) *In silico* prediction of best flyers peptides

When no experimental data are available at all
Prediction algorithms can be used to predict the best peptides to be used as targets

Prediction tools (physico-chemical properties)

- ESP predictor *(Fusaro V. et al., Nature Biotechnology, 2009)*
- *http://www.broadinstitute.org/cancer/software/genepattern/esppredictor*
- PeptideSieve *(Mallick P. et al., Nature Biotechnology, 2007)*
- *http://tools.proteomecenter.org/wiki/index.php?title=Software%3APeptideSieve*
- PepFly *(Sanders W. et al., BMC Bioinformatics, 2007)*
- *http://www.mybiosoftware.com/pepfly-peptide-flyability-prediction.html*
- …

71

Available proteomic data

Some peptides have better ionization / fragmentation than others.
To select the best peptides for targeted quantification of a protein, use preferentially:

1) Experimental proteomic data already acquired on your samples of interest

2) Experimental proteomic data on other samples in which your protein(s) of interest has/have been identified

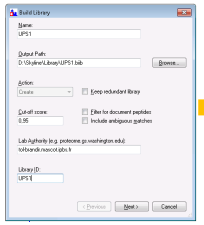
3) Peptides repositories available online

Generation of spectral libraries in Skyline

Skyline

72

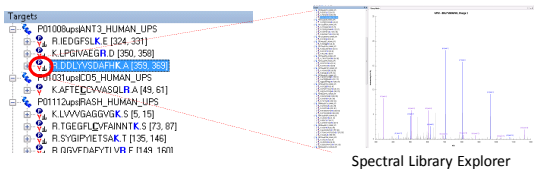
Spectral library in skyline



Build a spectral library from database search results of your .raw files

=>.dat (Mascot)
.msf (Sequest)
.omx (OMSSA)
.pepXML (Xtandem!)
.mzIdentML

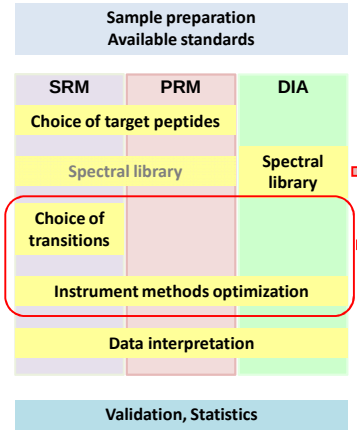
New: Import validated spectral library from Proline



Spectral Library Explorer

Spectral libraries from data repositories can be directly added in Skyline

Assay development workflow



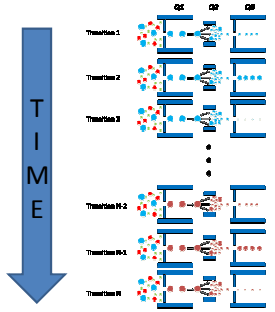
Spectral libraries are useful for all targeted strategies

Choice of transitions must be made *a priori* in SRM but *a posteriori* in PRM and DIA modes

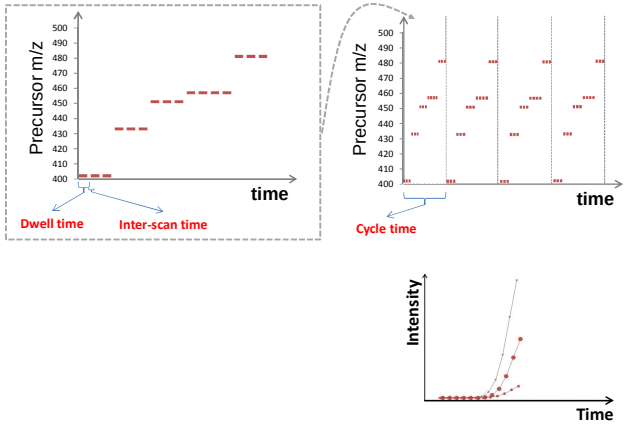
Method parameters setup : Dwell time / Cycle time / Inter-scan time / Time-Scheduling

SRM signal acquisition

During a SRM experiment:
→ Each transition is measured during a limited time (**dwell time**).
→ The mass spectrometer monitors all the **transitions sequentially**.
→ The time to measure all transitions is called the **cycle time**.

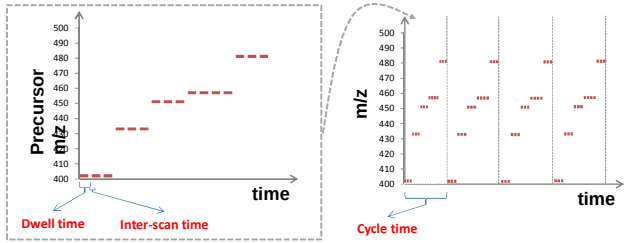


SRM signal acquisition



77

SRM signal acquisition



Dwell time: Time spent measuring a given transition.
Between 5ms and 100ms

Cycle time: Time spent measuring all transitions.
Time necessary to monitor the complete list of transitions.
Between 1.5s and 3s

Inter-scan time: Time necessary to switch to another transition.
Time necessary to change the voltages to monitor another transition.
Between 1 and 5ms

78

Cycle time calculation

$$\textbf{\textit{Cycle time}} = \textbf{\textit{Number of transitions}} \times (\textbf{\textit{Dwell time}} + \textit{interscan time})$$

- Has to be as short as possible
- Is fixed for each MS analysis
- Has to be as long as possible
- Is fixed

79

SRM method optimization

SRM experiments are optimized when these 3 parameters are in equilibrium

Sensitivity

Dwell time:
Has to be as long as possible

Between 5ms and 100ms

Accuracy

Cycle time:
Has to be as short as possible

Conditions the number of points measured per chromatographic peak

Between 1.5s and 3s

Multiplexing

Number of transitions:
Proteins, peptides, L/H…

3 to 6 transitions per peptide
4 to 5 peptides per protein
~1000 transitions per run

80

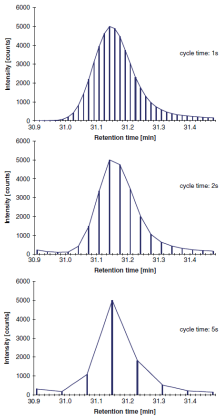
Effets of the cycle time on the accuracy

→ The longer the cycle time, the lower the number of points per chromatographic peak.

→ Quantification by SRM is done using the Area under the curve of the peak (or the height of the peak).

→ A minimum of 10 points per peak is necessary for good quantification.

Do not forget to optimize the chromatography!



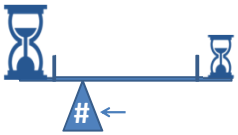
S. Gallien, E. Duriez, B. Domon, J Mass Spectrom 46, 298 (Mar, 2011).

SRM method optimization

Optimize the number of followed transitions

Sensitivity

Dwell time:
Has to be as long as possible



Accuracy

Cycle time:
Has to be as short as possible

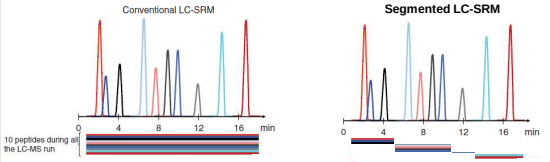
Number of points per chromatographic peak

Multiplexing

Number of transitions at a given time

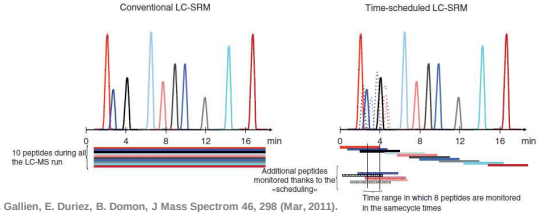
How to optimize dwell, cycle time?

- **By segments**



-Good when segments are well defined (Not always)
- Overlaps can be problematic

- **Scheduled/Dynamic SRM**

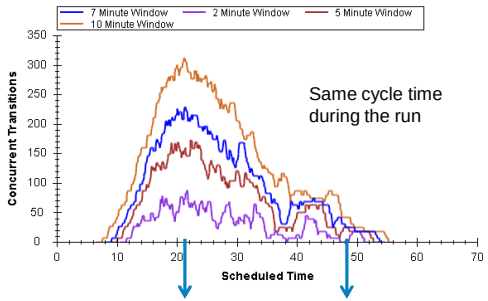


For each peptide the minimal information necessary for a SRM experiment is :
-Precursor m/z
-Product m/z
-Retention time

S. Gallien, E. Duriez, B. Domon, J Mass Spectrom 46, 298 (Mar, 2011).

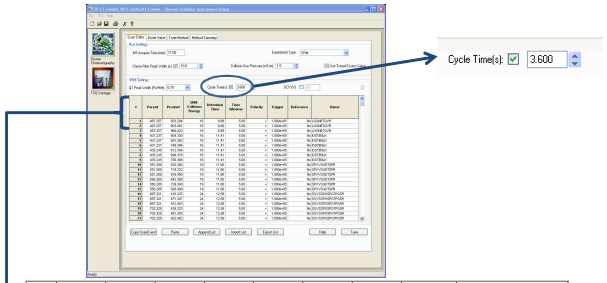
Scheduled SRM

For a given cycle time, the dwell time will change during the run.
This maximizes/optimizes the dwell time for each transition.



SRM method

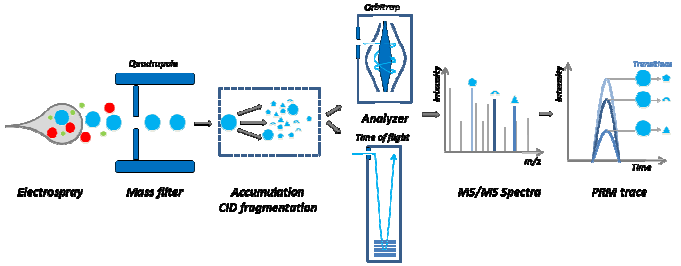
On the mass spectrometer the cycle time is fixed



#	Parent	Product	SRM Collision Energy	Retention Time	Time Window	Polarity	Trigger	Reference	Name
1	487.257	503.294	18	8.88	5.00	+	1.000e+05	No	LGGNETQIVR
2	487.257	803.401	18	8.88	5.00	+	1.000e+05	No	LGGNETQIVR
3	487.257	860.422	18	8.88	5.00	+	1.000e+05	No	LGGNETQIVR

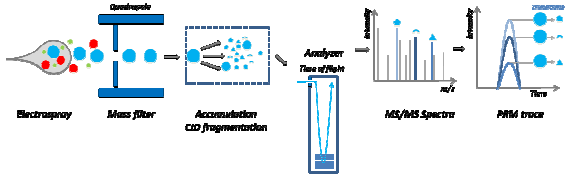
85

PRM signal acquisition



86

PRM signal acquisition on Q-TOF

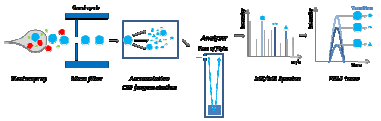


For Q-TOF instruments (AB Sciex):

- The cycle time is kept constant
- Accumutation-time dependent on the number of concurrent ions to be fragmented
- Resolution is not related to scanning time
- All principles seen earlier for scheduled SRM are valid

87

PRM signal acquisition on Q-TOF



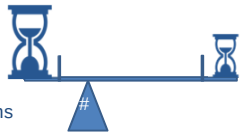
For Q-TOF instruments (AB Sciex):

- The cycle time is kept constant
- All principles seen earlier are valid

Sensitivity

Accumulation time:
Has to be as long as possible

Between 50ms and 100ms



Accuracy

Cycle time:
Has to be as short as possible

Number of points per chromatographic peak

Between 1.5s and 3s

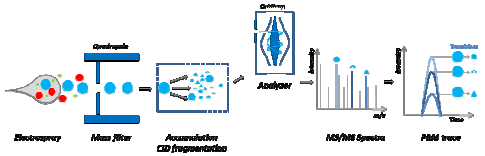
Multiplexing

Number of precursors at a given time

4 to 5 peptides per protein
40-60 concurrent precursors max

88

PRM signal acquisition on Q-Orbitrap

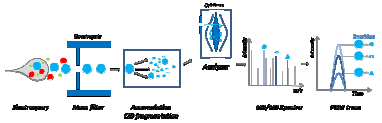


In FT-based instruments resolution is directly proportional to the transient Length

Resolving Power at m/z 200	Resolving Power *at m/z 400*	Transient length (ms)	Approximate scan speed (Hz)
17,500	*12,500*	64	13
35,000	*25,000*	128	7
70,000	*50,000*	256	3
140,000	*100,000*	512	1.5

89

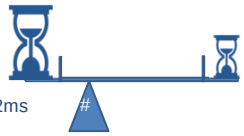
PRM signal acquisition on Q-Orbitrap



Sensitivity
Resolution

Transient time:
Fixed (Set by the user)

Between 64ms and 512ms



Accuracy

Cycle time:
Has to be as short as possible

Number of points per chromatographic peak

Between 1.5s and 3s

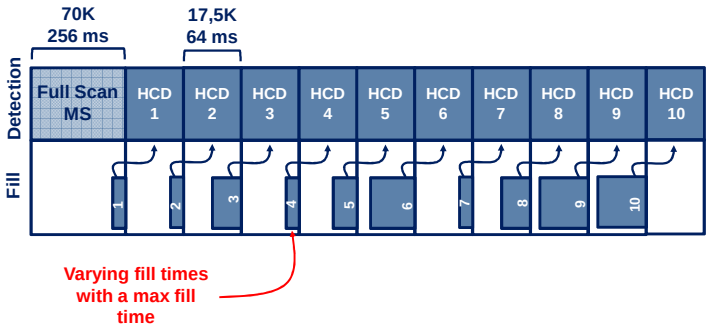
Multiplexing

Number of precursors at a given time

4 to 5 peptides per protein
8-32 concurrent precursors max

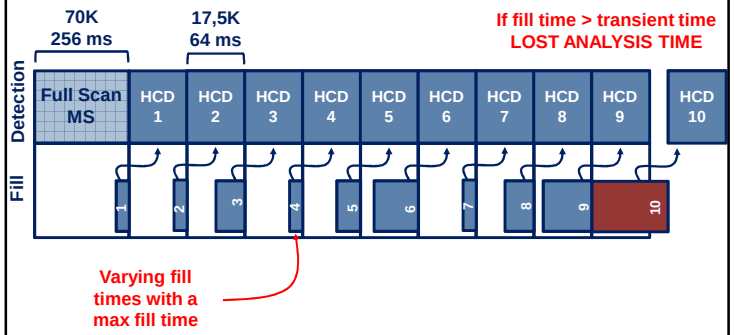
90

PRM signal acquisition on Q-Orbitrap



91

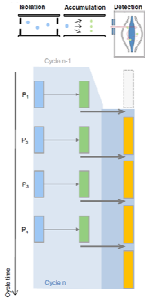
PRM signal acquisition on Q-Orbitrap



92

PRM signal acquisition on Q-Orbitrap

A. Simplex mode



A

if Max. Fill Time < Transient time

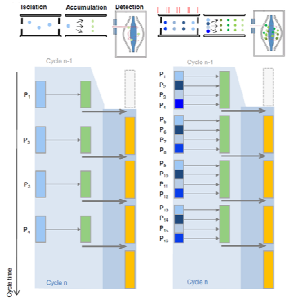
Cycle time = Number of precursors × (Transient time)

Changes throughout the gradient | Changes throughout the gradient | Fixed value (Defined by the resolving power used)

Gallien S, Bourmaud A, Kim SY, Domon B: Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics 2014, 100:147-159.*

PRM signal acquisition on Q-Orbitrap

A. Simplex mode B. Multiplex mode



A

if Max. Fill Time < Transient time

Cycle time = Number of precursors × (Transient time)

Changes throughout the gradient | Changes throughout the gradient | Fixed value (Defined by the resolving power used)

B

$$\text{if Max. Fill Time} < \frac{\text{Transient time}}{\text{Multiplexing degree}}$$

$$\text{Cycle time} = \text{Number of precursors} \times \left(\frac{\text{Transient time}}{\text{Multiplexing degree}}\right)$$

Gallien S, Bourmaud A, Kim SY, Domon B: Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics 2014, 100:147-159.*

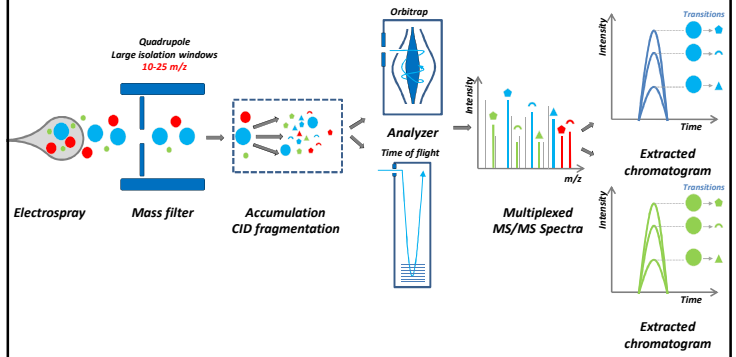
PRM signal acquisition on Q-Orbitrap

Multiplexing degree	1	1	1	4	8
Mode of operation[a]	Sx	Sx	Sx	Mx	Mx
Maximum fill time/peptide [ms]	250	120	60	30	15
'Total' maximum fill time [ms]	250	120	60	120	120
Resolving power[b]	70 k	35 k	18 k	35 k	35 k
Transient time [ms]	256	128	64	128	128
Cycle time [s]	2	2	2	2	2
Isolation window/peptide [m/z unit]	2	2	2	2	2
'Total' isolation window [m/z unit]	2	2	2	8	16
Number peptides/cycle	8	16	32	64	128

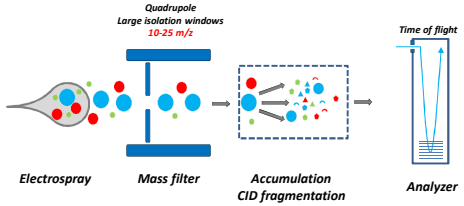
Multiplexing increases

- Selectivity decreases
- Sensitivity decreases
- Dynamic range can also decrease due to high-abundant co-isolated peptides

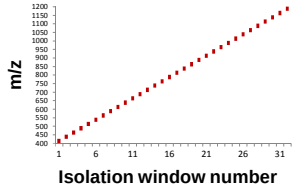
Data-Independent Acquisition



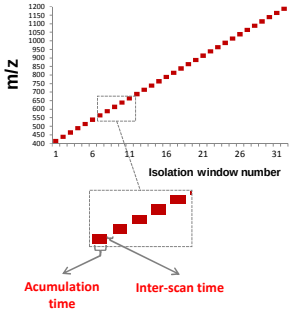
SWATH-MS

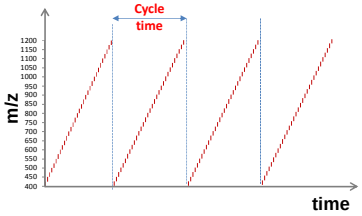


- High resolution
- High scan rate



SWATH-MS





DIA method optimization

Selectivity

Window size:
Has to be as small as possible

Sensitivity

Accumulation time:
Has to be as long as possible

Between 50ms and 100ms

Accuracy

Cycle time:
Has to be as short as possible

Conditions the number of points measured per chromatographic peak

Between 1.5s and 3s

Multiplexing

Number of windows:
m/z range covered
Depends of:

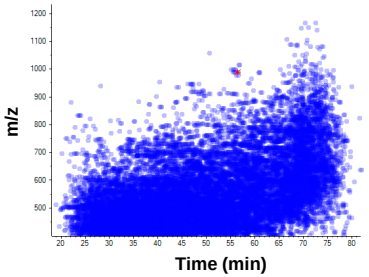
- sample
- instrument

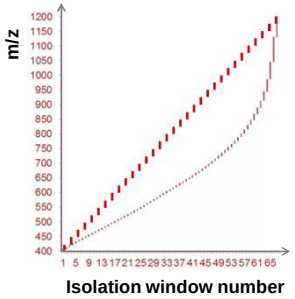
Number of windows	32
Cycle time (s)	3
Acc. Time (MS1) (ms)	150
Acc. Time (MS2)	89

99

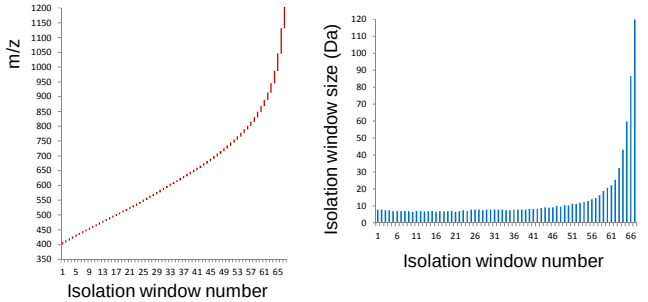
Peptide distribution

Yeast sample
80 min gradient

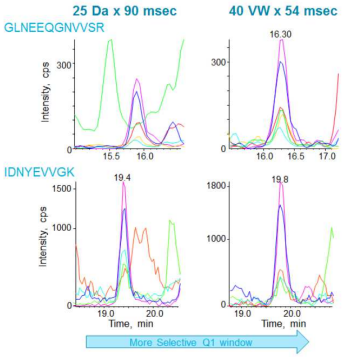




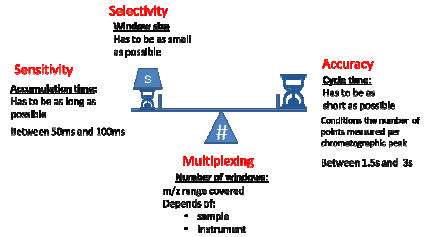
Variable windows SWATH



Increase selectivity



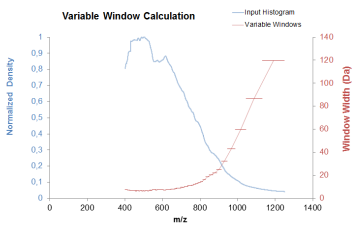
How many windows?



400-1250 m/z

Number of windows	32	50	65	75	100
Cycle time (s)	3	3	3	3	3
Acc. Time (MS1) (ms)	150	150	150	150	150
Acc. Time (MS2)	89	57	44	38	29

How to chose the window width



SWATH Variable Window Assay Controls

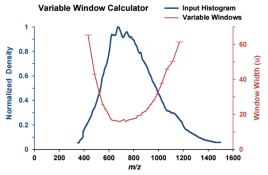
Target number of windows:	67	(actual # may be less depending on min window width setting)
Lower m/z limit:	400	(min 200)
Upper m/z limit:	1250	(max 1250 for TripleTOF 5600+, max 2250 for TripleTOF 6600)
Round bin edges to x figures:	1	(1 figure past decimal recommended)
Window overlap (Da)	1,0	(1 Da overlap recommended)
Minimum window width (Da)	3	
CES	10	

How to chose the window width

Journal of proteome research
Article
pubs.acs.org/jpr

The Use of Variable Q1 Isolation Windows Improves Selectivity in LC–SWATH–MS Acquisition

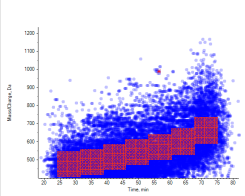
Ying Zhang,[†,⊥] Aivett Bilbao,[‡,⊥,▽] Tobias Bruderer,[†] Jeremy Luban,[∥] Caterina Strambio-De-Castillia,[∥] Frédérique Lisacek,[‡,§] Gérard Hopfgartner,[*,†] and Emmanuel Varesio[*,†]

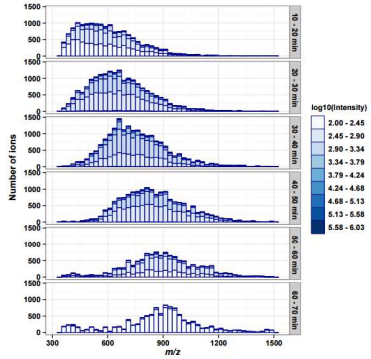


Different distribution than our sample

Peptide distribution depends on the sample!!

Towards Scheduled SWATH?





Not yet implemented in instrument software and data treatment software

Choice of transitions / Collision energy optimization

Principle and objective

- Matrix mixture = mimic all possible interferences
- Used as reference sample all over the analyses

Before selection

After selection

Delete transitions :
Low intensity
Non specific

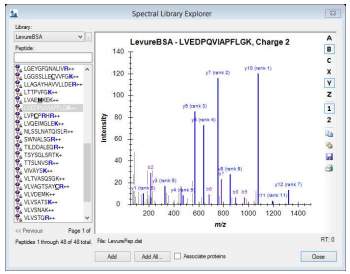
Objective :
Test a maximum of transitions, min 3 transitions conserved

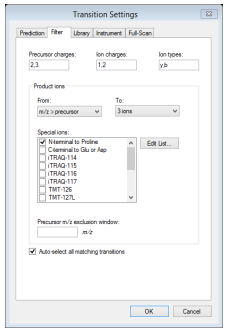
How to choose transitions to test

Spectral library from MS/MS results
(shotgun DDA datasets or MS/MS data of heavy labeled reference peptides)

And / Or

Theoretical fragments from skyline



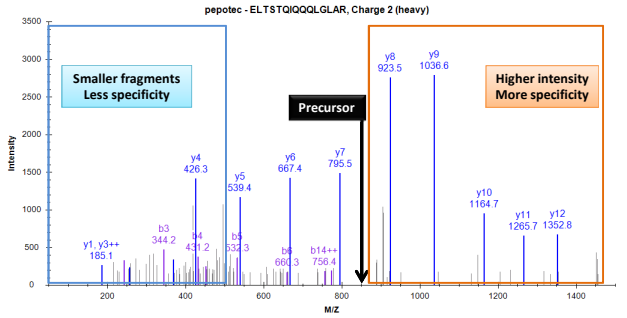


How to choose transitions to test

Some general rules :

More than 4 amino acids

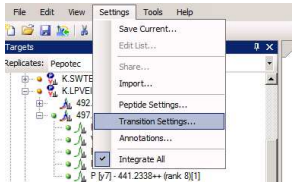
Fragment m/z> Precursor m/z

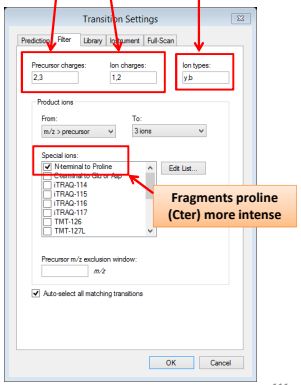


Theoretical choice

State of charge
Precursor - Fragment

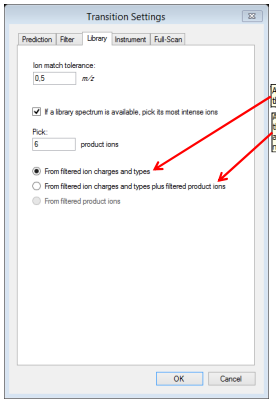
Ion type
y, b...





Fragments proline (Cter) more intense

Selection from spectral library



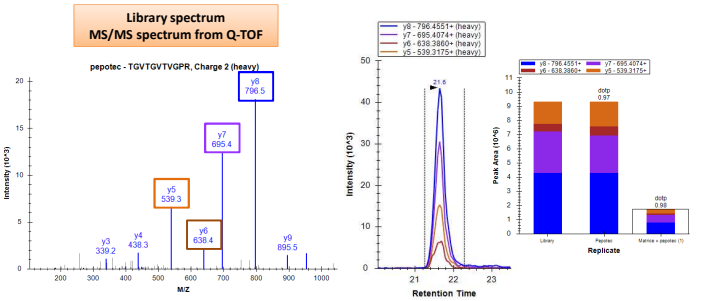
- Method export from skyline to MS software
- Standard sample injection :
  - Matrix + heavy peptides

  or Heavy peptides alone
- Work in unscheduled mode to determine retention time

Transition selection: DotProd

Good transition : Intense + Specific

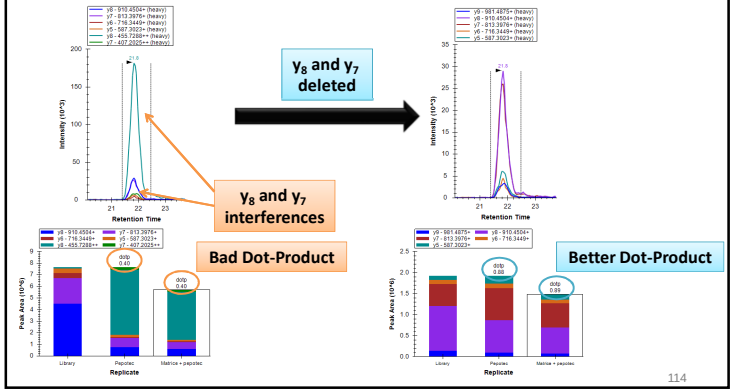
dotp (dot product) : correlation between the peak intensities of the transitions for the peptide of interest and the library spectrum for the same peptide.



Prakash,A. *et al.* J Proteome.Res. 2009
Sherwood,C.A. *et al. J. Proteome. Res.* 2009

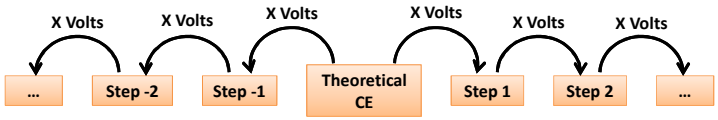
Transition selection: DotProd

Interference example



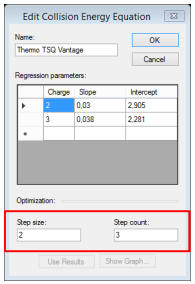
Collision energy optimisation

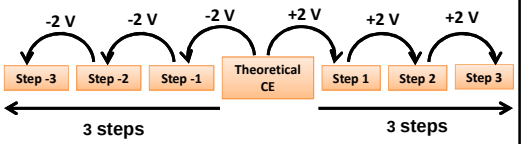
- **Improve fragmentation to obtain best fragments intensity in SRM**
- In Skyline: theoretical equation, a function of charge state and m/z
- Optimization : test of several CE around theoretical CE



Final number of transitions = Number of transitions x number of tested CE

Step size – Step count





Sample: UPS1

48 proteins

106 peptides

318 transitions

318 transitions x 7 CE

= 2226 transitions

Max concurrent transitions

Unscheduled mode

Cycle time = 3,5 s

318 transitions x 7 CE = 2226 transitions

Single method

Dwell time = 3,5 ÷ 2226 = 1,6 ms

Too short to have enough sensibility : Several methods needed

Several methods

Dwell time min = 25 ms

Max concurrent transitions = 3,5 ÷ 0,025 = 140

Number of methods = 2226 ÷ 140 = 15,9 ➔ 16 methods

Too many methods : Scheduling needed

Max concurrent transitions

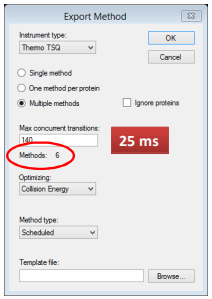
Scheduled mode

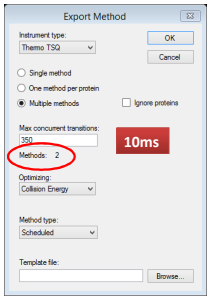
Cycle time = 3,5 s

318 transitions x 7 CE = 2226 transitions

Dwell time min = 25 ms

Max concurrent transitions = 3,5 ÷ 0,025 = 140





Max concurrent transitions

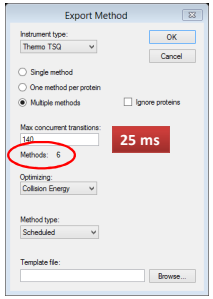
Scheduled mode

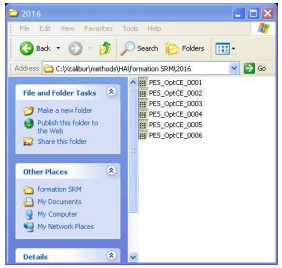
Cycle time = 3,5 s

318 transitions x 7 CE = 2226 transitions

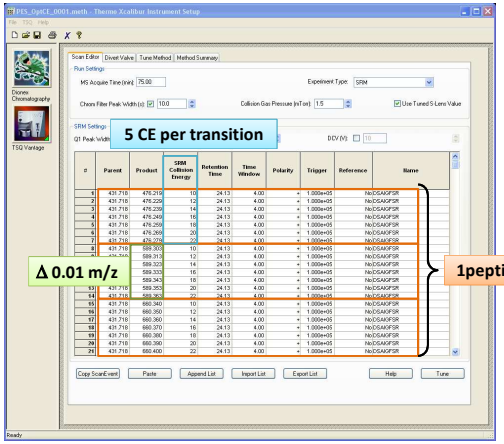
Dwell time min = 25 ms

Max concurrent transitions = 3,5 ÷ 0,025 = 140

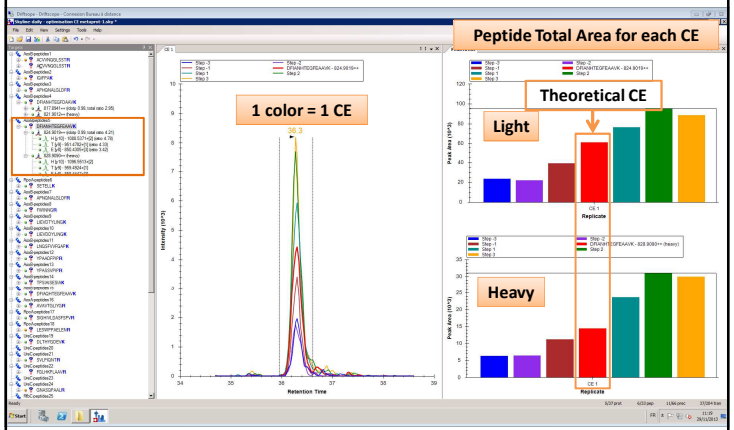




Methods export

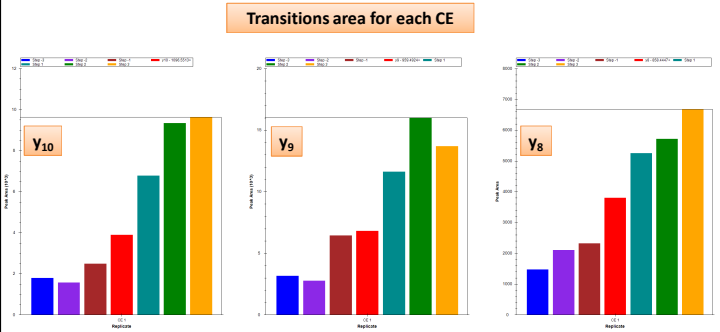


CE visualization in Skyline



121

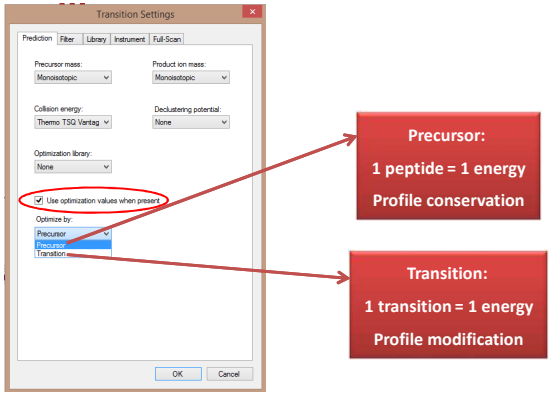
CE visualization in Skyline



It is worth optimizing collision energy for best sensitivity !!!

122

Precursor vs transition



Precursor:
1 peptide = 1 energy
Profile conservation

Transition:
1 transition = 1 energy
Profile modification

123

Conclusion

Transitions :

- Test several transitions in matrix
- Choose 3 transitions minimum
- Prefer y- ions
- Fragments with more than 4 amino acids
- m/z fragment > m/z precursor

Collision Energy

- Use equation
- If possible: optimization on heavy peptides for increased sensitivity
- Same energy for one peptide or one energy per transition

124

Use of retention time reference peptides

Sources of RT variation

- Peptide physico-chemical properties
  -AA sequence, PTMs,…
- LC system
  -column, solvents, pumps, mixers, capillaries…
- Matrix effects
  -Pure solvent, plasma, amount of sample…

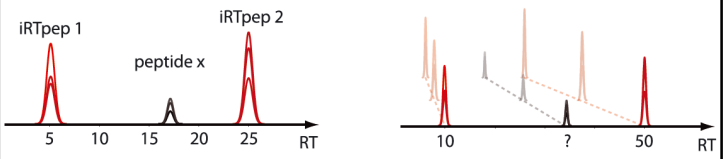
How to determine a peptide's retention time

Methods

	Direct approach	Databases	In silico calculation SSRCalc
Description	- Perform an **Uncheduled SRM** experiment - Determine RT of target peptides	- **Look for RT values in databases** or previous experiments (LC-MS/MS)	- **Hydrophobicity index** (HI) from **peptide sequence** - Measure RTs from **calibrant peptides** - Linear regression HI =ax(RT)+b - Calculate RT=(HI-b)/a for all target peptides
Pro	- Straightforward - Experimental data	- Experimental data	- **Only runs to analyze calibrants are needed**
Con	- **Time consuming** - **Many SRM methods** needed - Needs to be repeated if conditions change - **Pre-runs** required before each experiment	- Time consuming - **Not directly applicable** - Needs to be adjusted if conditions change - **Pre-runs** required before each experiment	- **Not accurate RT** prediction - **Large RT windows** required - Modifications are not taken into account

iRT Prediction

Combines experimental measurement and in-silico prediction



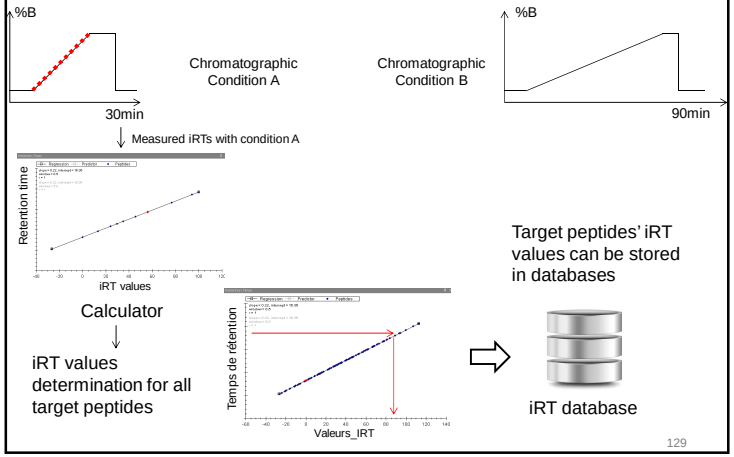
- **Measure RT for calibrants (iRT) and target peptides**
- **Obtain iRT value** for each target peptide
  - Dimensionless
  - Relative to calibrant peptides
  - Independent of LC gradient
- This has to be done only once
- Information can be stored in databases

- Measure calibrant (iRT) peptides RT
- Linear regression iRT =ax(RT)+b
- Calculate RT= (iRT-b)/a for all tageted peptides

→ Fully implemented into Skyline

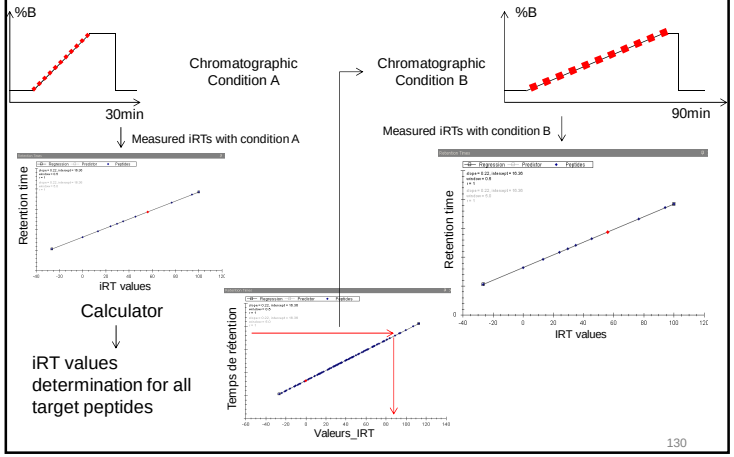
C. Escher *et al.*, *Proteomics* 12, 1111 (Apr, 2012)

iRT Prediction



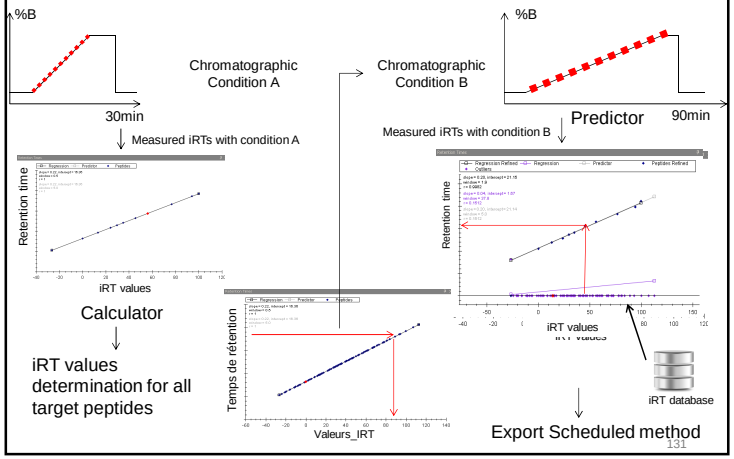
129

iRT Prediction



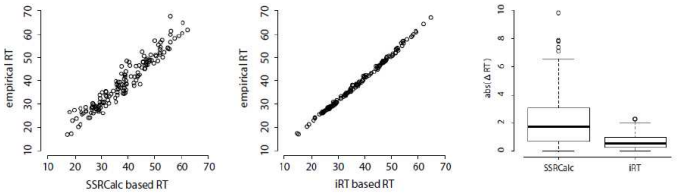
130

iRT Prediction



131

iRT avantages



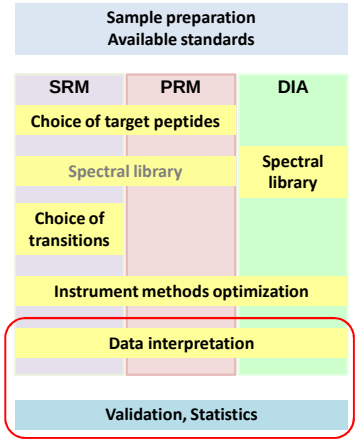
→ Time-saving approach
→ High multiplexing of RT prediction
→ No pre-runs required (except iRT standards)
→ PTMs are taken into account

Essential for any type of targeted method (SRM, PRM and DIA)

C. Escher *et al.*, *Proteomics* 12, 1111 (Apr, 2012)

132

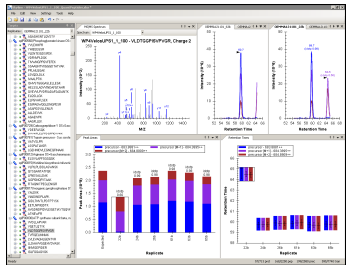
Assay development workflow



Data Interpretation

Skyline is an essential tool for targeted data interpretation

-Multiple views of replicates
-Easy data checking: manual verification is possible, in a fast and efficient way
-Visualisation of interferences
-Flexible and rich export templates



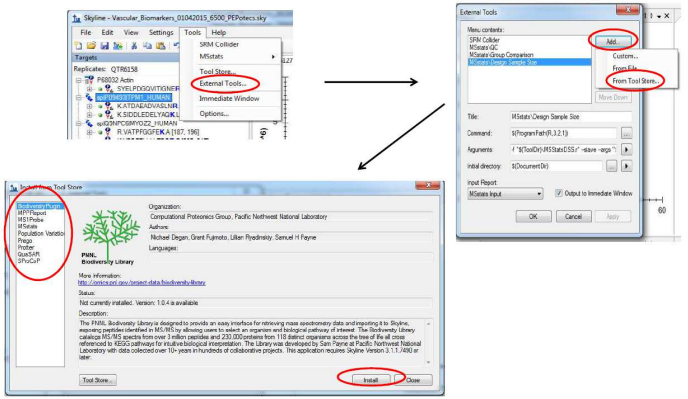
MacLean B. et al., Bioinformatics 2010

Data Interpretation

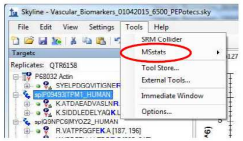
Skyline is an essential tool for targeted data interpretation

MacLean B. et al., Bioinformatics 2010

External tools



MSstats



What is MSstats?

→ Obtention of statistical and visual data

→ Quantification of proteins in biological samples

→ Design of future experiments

Bioinformatics. 2014 Sep 1;30(17):2524-6. doi: 10.1093/bioinformatics/btu305. Epub 2014 May 2.

MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments.

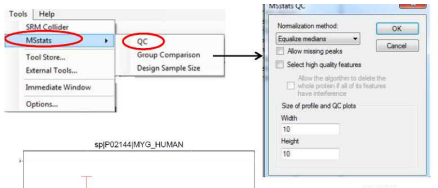
Choi M[1], Chang CY[1], Clough T[1], Broudy D[1], Killeen T[1], MacLean B[1], Vitek O[2].

Author information

Abstract

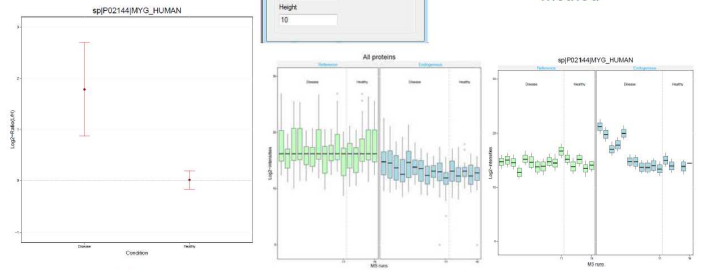
MSstats is an R package for statistical relative quantification of proteins and peptides in mass spectrometry-based proteomics. Version 2.0 of MSstats supports label-free and label-based experimental workflows and data-dependent, targeted and data-independent spectral acquisition. It takes as input identified and quantified spectral peaks, and outputs a list of differentially abundant peptides or proteins, or summaries of peptide or protein relative abundance. MSstats relies on a flexible family of linear mixed models.

MSstats



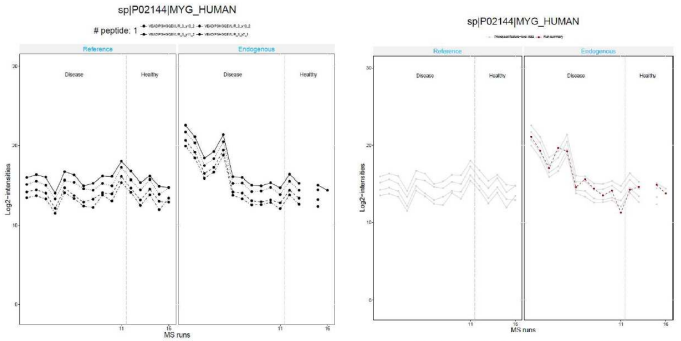
Choose your parameters→Submit Run of R package automatically

Several files are saved in the file where you stored the skyline method



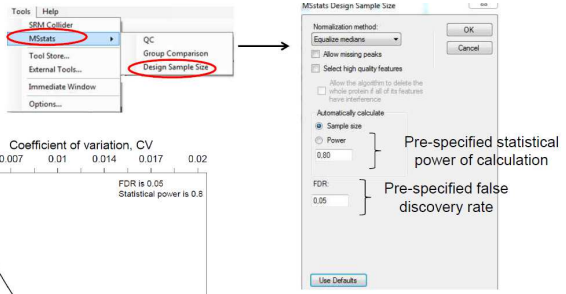
Condition Plot **QC Plot**

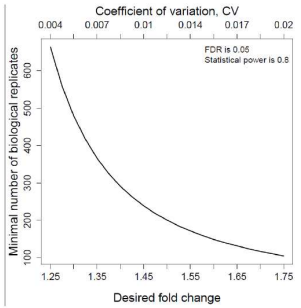
MSstats



Profile Plot **Profile Plot Summarization**

MSstats



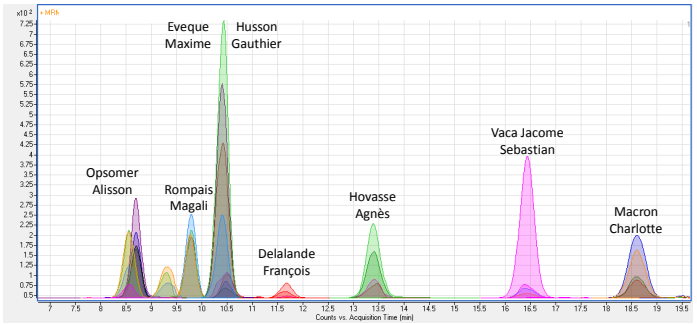


Pre-specified statistical power of calculation

Pre-specified false discovery rate

Design sample size: The minimum number of samples required to obtain the desired protein abundance log fold change → Reliable statistics

Thanks!



WP3 of the ProFI:
Alexandre Stella, Luc Garrigues, Anne Gonzalez,
Mohammed Benama, Michel Jaquinot, Floriane Pailleux

141