Lessons Learned the Hard Way:
Acquiring and Analyzing DIA Proteomics Data with Orbitraps

Brian C. Searle, Lindsay K. Pino, Seth C. Just, and Michael J. MacCoss

ISB Penn University of Pennsylvania Scaffold Proteome Software University of Washington

bsearle@systemsbiology.org
@briansearle

Hi, my name is Brian Searle and I'd like to tell you about some lessons learned the hard way on acquiring and analyzing data independent acquisition (or DIA) data with Orbitraps. To begin, this talk is creative commons open-source licensed, so feel free to email me for the slides and remix them into your own presentations with attribution. If you have follow-up questions, please either email me directly or tweet at my twitter handle. I'd like to start out with my acknowledgements first: this work couldn't have been done without Lindsay, Seth, and Mike. I thank them for all the work they've put in.

Speaker Disclosures
Brian C. Searle

Relevant financial relationships:

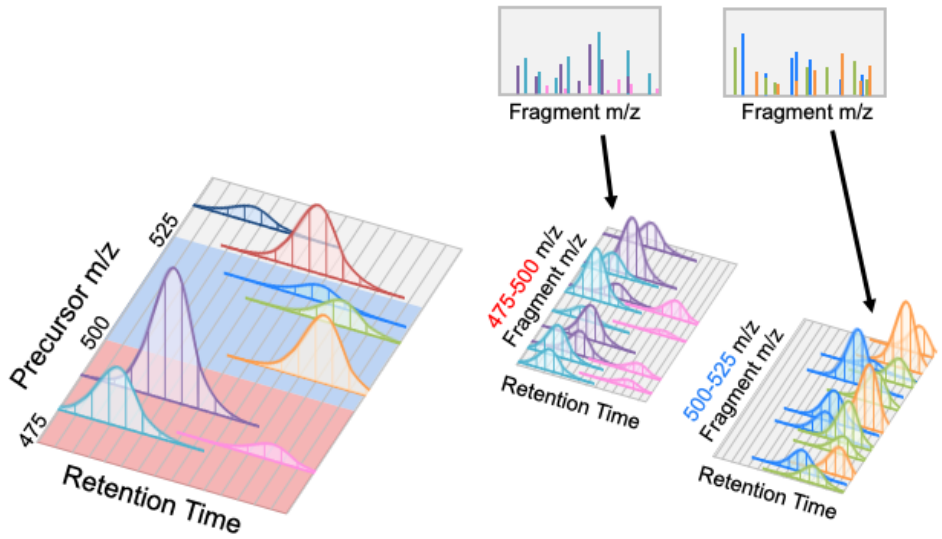
- Employed as a fellow at the Institute for Systems Biology

- Founder and shareholder at Proteome Software

Research funding sources:

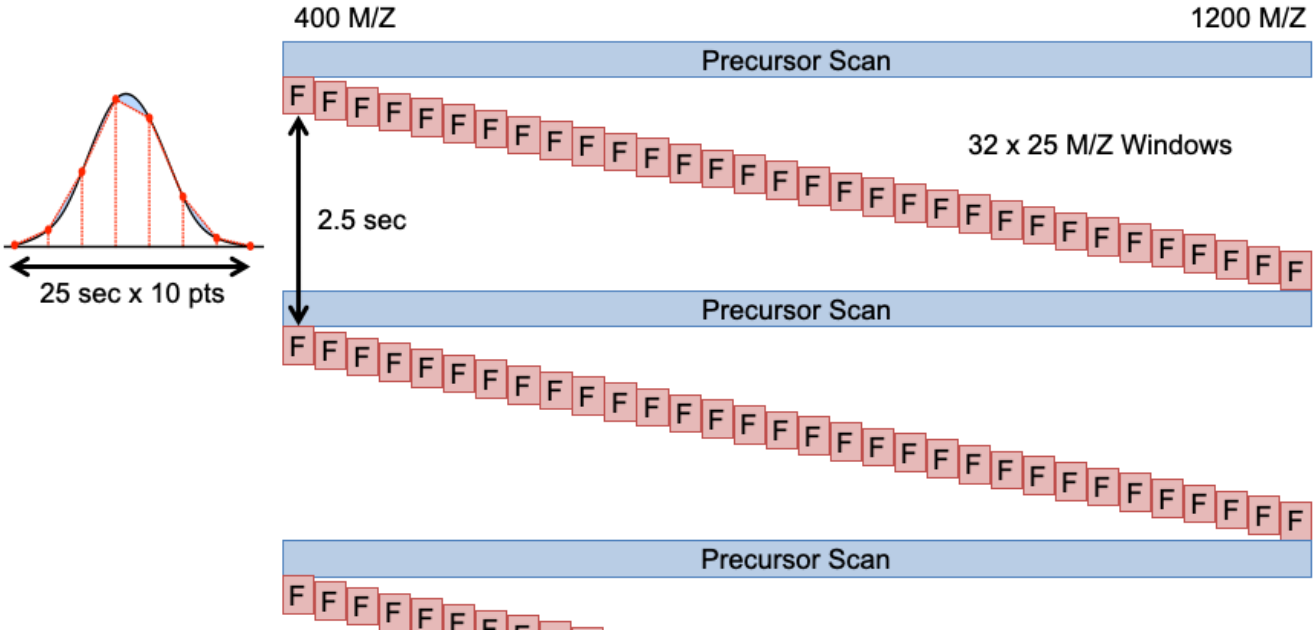
Before I get started, the ISB asked me to show you this speaker disclosure. It just says that I am employed at the ISB, but I'm also a founder and shareholder of Proteome Software.

Full proteome coverage at a cost:
increased complexity due to multiplexing



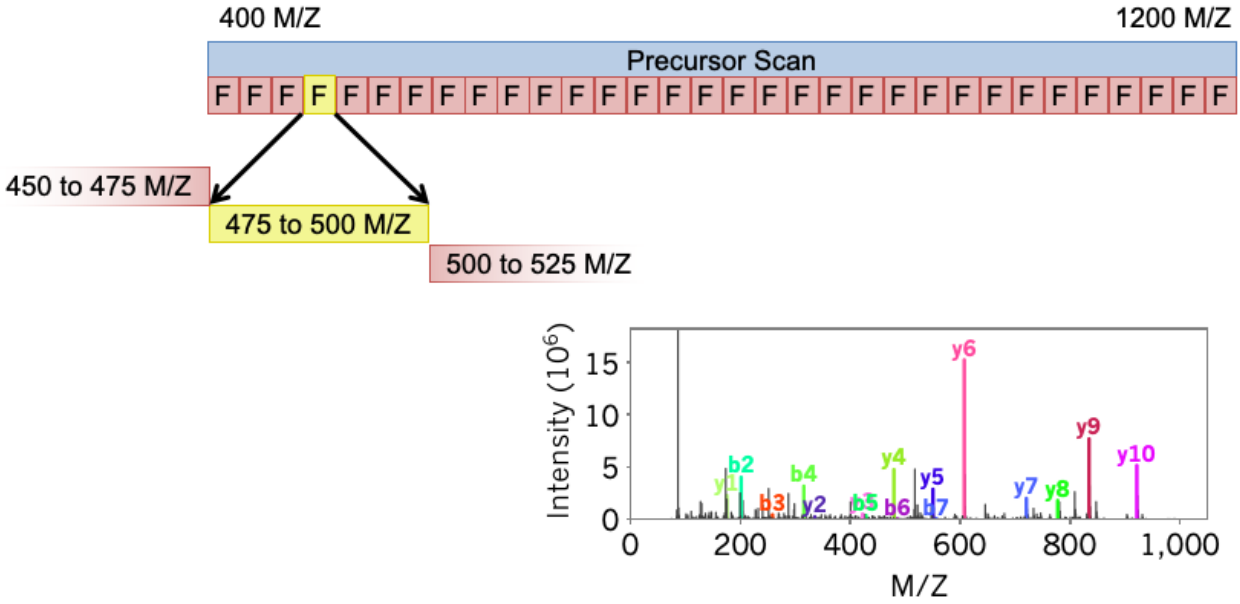
We've heard some about DIA already in this session, but in case you've just come in or need a refresher: DIA is a multiplexing acquisition strategy that repeatably measures peptides using fixed precursor isolation windows. For example, in this schematic there are 3 peptide signals with precursors found between 475 and 500 m/z. The peptides in this range are selected by a quadropole and co-fragmented together, creating complex MS/MS.

Comprehensive data acquisition



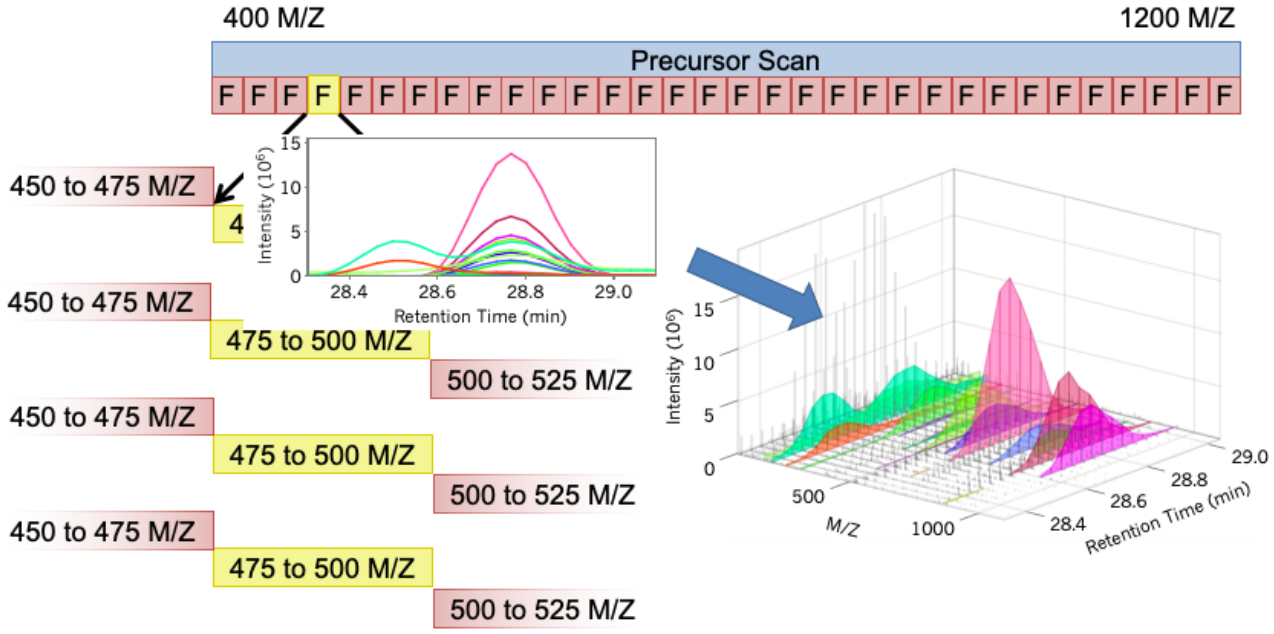
In this strategy, generally people acquire a precursor spectrum, followed by sequential fixed window MS/MS. Each window is repeatedly measured every 2-3 seconds, ensuring enough points across the peak to enable accurate peptide quantification.

LSGGLGAGSCR +2H (489.2 M/Z)



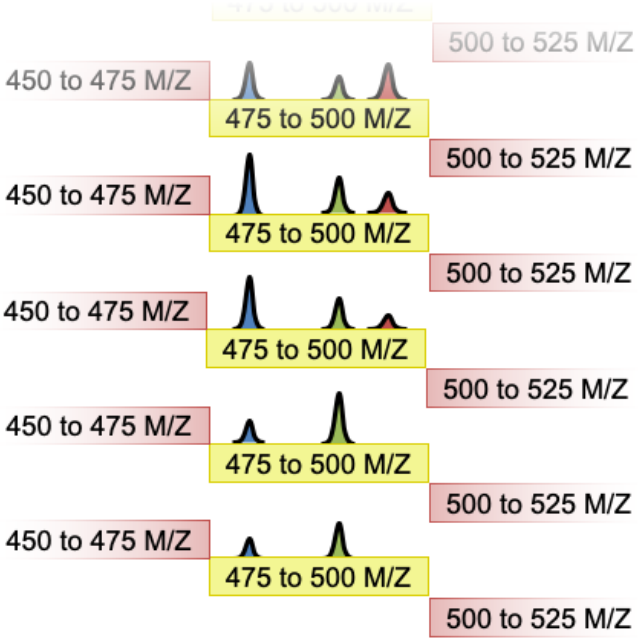
Let's look at a peptide in the 475 to 500 m/z window. At the apex, we might see a spectrum like this

LSGGLGAGSCR +2H (489.2 M/Z)

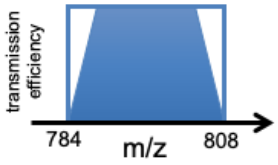


But because we have multiple measurements covering the peptide, we actually acquire 3-dimensional fragment-level data. We visualize these fragment ions as chromatograms.

Complications with DIA…



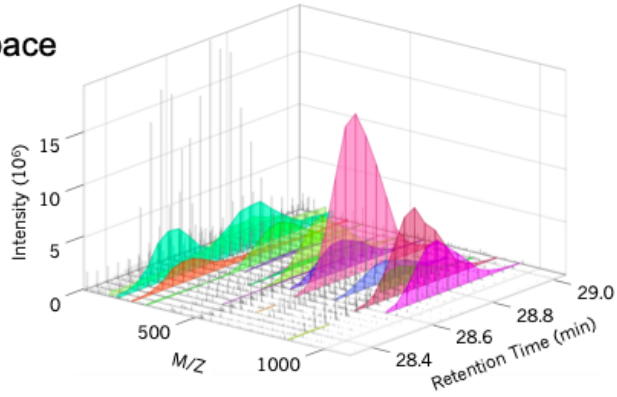
- Ions are not centered or collision energy optimized
- Usually there are multiple peptides
- Edge effects can suppress signal



Comprehensive quantification of all peptides? Possibly, but there are some complications. First, peptides can't be collision energy optimized based on their charge state. As I said earlier, multiple peptides are multiplexed, which complicates spectra. Finally, quadropoles don't have perfect transmission and there are edge effects that can suppress signal near the window boundaries.

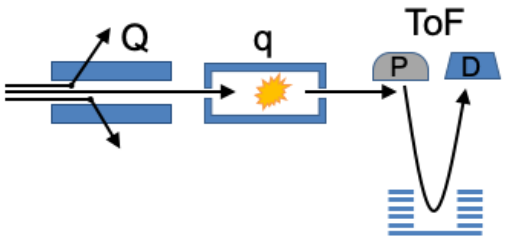
Contents

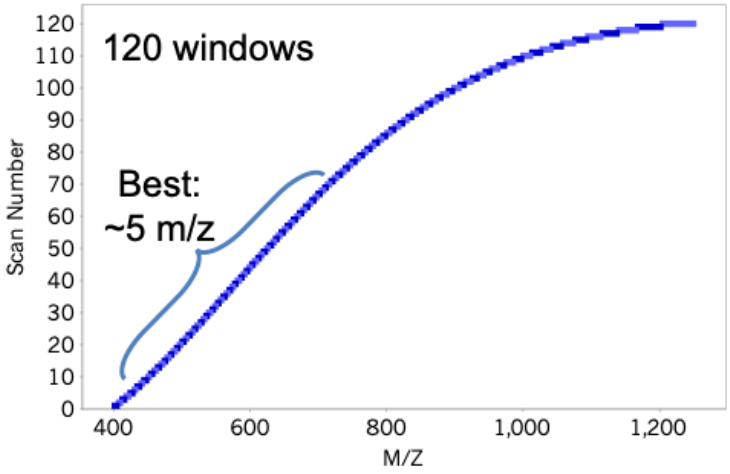
- How is DIA different for ToFs and Orbitraps?
- Re-designing DIA to take advantage of Orbitraps
  - Narrow precursor range
  - Stagger windows to under-sample search space
  - Take advantage of “forbidden zones”
- Generating DIA-only libraries
- Resources for DIA best practices



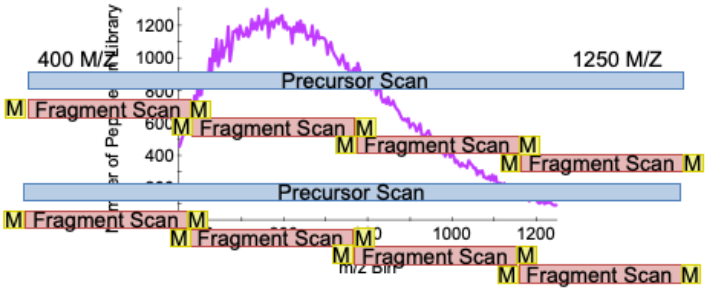
With that out of the way, I want to tell you about how DIA differs on ToF and Orbitrap instruments, and how to re-design DIA measurements to take advantage of Orbitrap features and hide their weaknesses. DIA search engines use peptide libraries, and in many circumstances, it is impossible or impractical to generate large DDA libraries. I’ll discuss how you can rapidly generate DIA-only libraries, and then end with some resources we’ve built to help you acquire high quality DIA datasets.

ToF acquisition



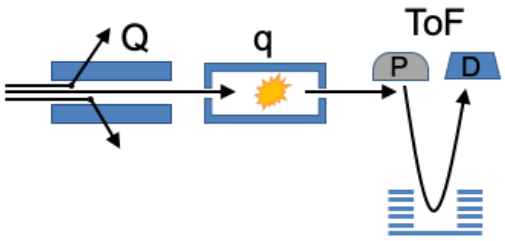


- Very fast! (1000s of pushes / sec)
- True profile scans
- Always see “signal” after averaging sufficient scans



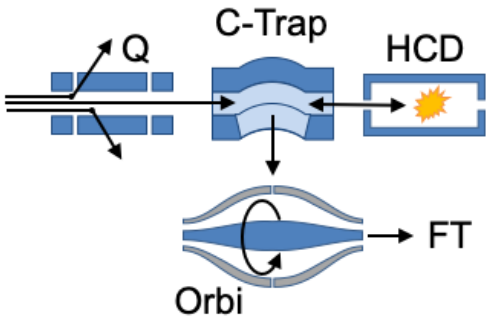
Modern DIA methods were originally developed and trademarked as SWATH on time-of-flight instruments. They took advantage of ToF features, such as very fast scan rates and after noticing that peptides tend to clump between 400 and 900 m/z, 32-fixed window DIA became variable-width DIA, where windows were narrower in this sweet spot and commensurately wider at higher or lower m/z. ToFs scan fast, so 32 windows became 40 became 60 became 120 with quite small 5 m/z windows for a large range of peptides. Because these windows were so small, many peptides suffered from edge effects due to incomplete quadrupole transmission. Considering the actual scans themselves, small margins were added to windows to compensate for quad isolation edge effects.

ToF acquisition



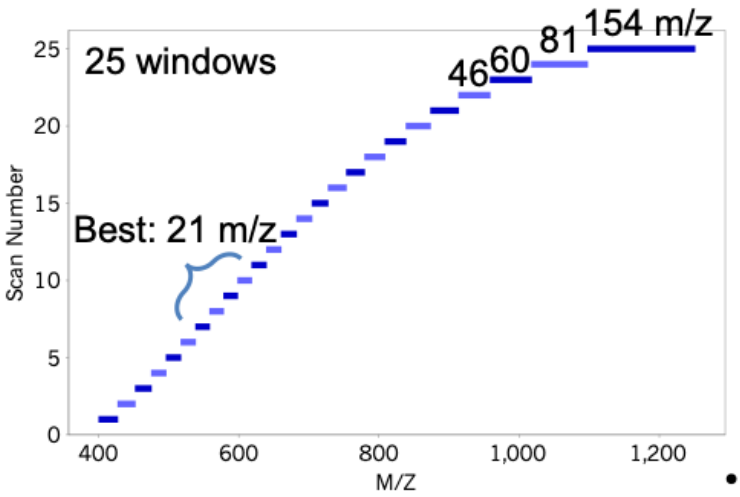
- Very fast! (1000s of pushes / sec)
- True profile scans
- Always see “signal” after averaging sufficient scans

Orbitrap acquisition



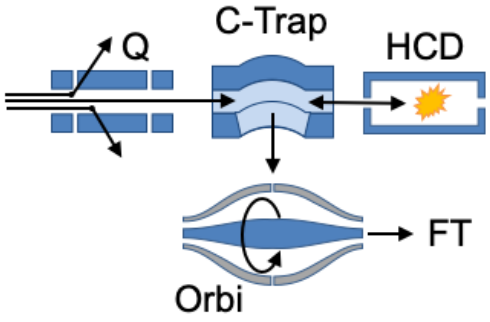
- Relatively slow (10-20 MS/MS / sec)

Now, lets compare that to Orbitrap instruments, which are comparably very slow, collecting MS/MS at 10 to 20 Hz.



10 Hz * 25 windows = 2.5 sec

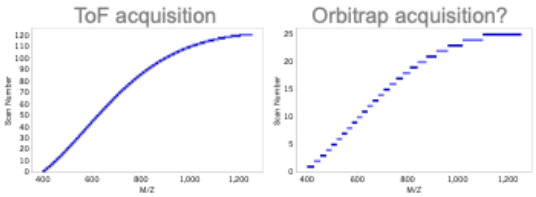
Orbitrap acquisition



- Relatively slow (10-20 MSMS/sec)
- Pseudo-profile from FT (built-in denoising)
- Segmented quad: flat(ish) transmission

If we applied that same DIA strategy, to achieve a 2.5 second duty cycle we'd be working with about 25 very-wide windows that are over 4 times larger. However, there are some advantages to Orbitraps that this strategy doesn't take advantage of. First, Orbitrap instruments use trapping devices to help use a higher proportion of the ion beam. Measurements are Fourier transformed, which helps filter noise. Additionally, since the release of the QE-plus in 2013, they've incorporated a segmented quad, which improves transmission efficiency across the window.

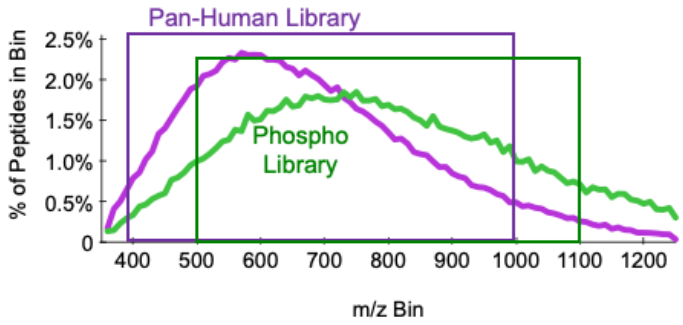
Contents

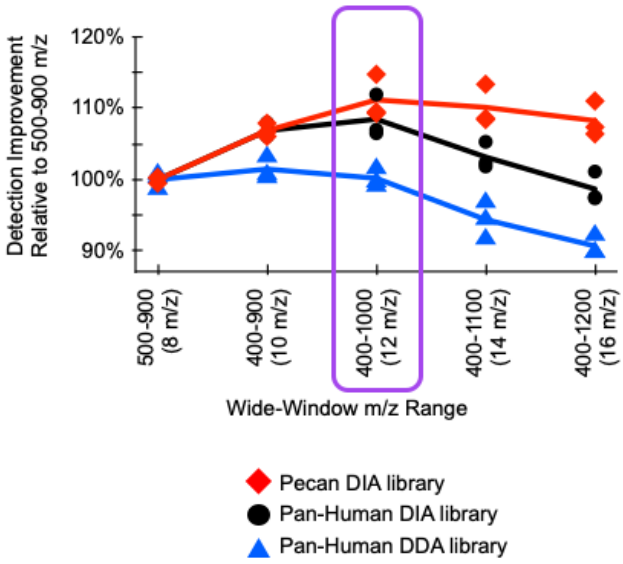


- How is DIA different for ToFs and Orbitraps?
- Re-designing DIA to take advantage of Orbitraps
  - Narrow precursor range
  - Stagger windows to under-sample search space
  - Take advantage of “forbidden zones”
- Generating DIA-only libraries
- Resources for DIA best practices

Let’s try to build a method from the ground up that uses these features to our advantage.

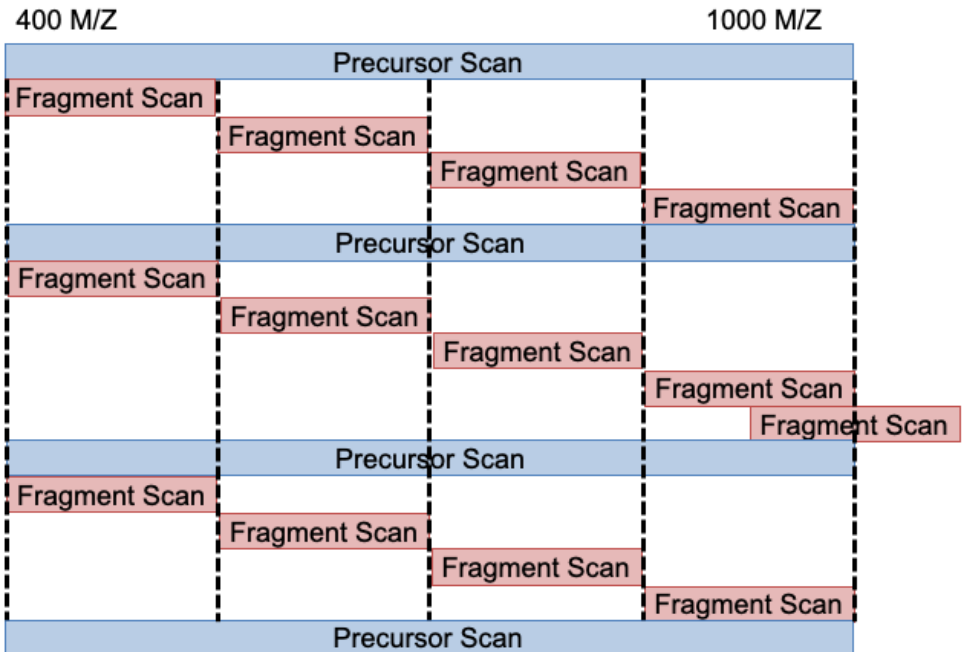
Why limit the precursor range?



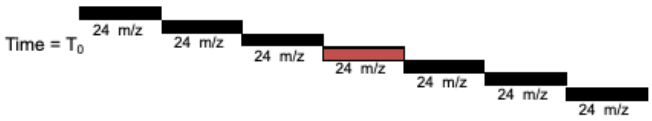


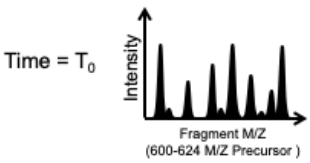
First, we're collecting too many scans. Rather than going up to 1250 m/z, we should take advantage of that plot from earlier that shows that most peptides fall in a narrower range, such as this 400-1000 range I'm showing here. Looking at HeLa, we ran replicate injections with five different windowing schemes with increasing window widths from 500-900 m/z to 400-1200 m/z. While widening the range helped to a point, as windows got much wider the detection rates dropped, no matter what library search strategy was used. We found that 400-1000 m/z was the optimal sweet spot, however, we only know this works for normal tryptic peptides. Other types of peptides, for example phosphopeptides, will likely have different optimal ranges, for example 500-1100 m/z.

Staggered windows uses under-sampling and signal reconstruction

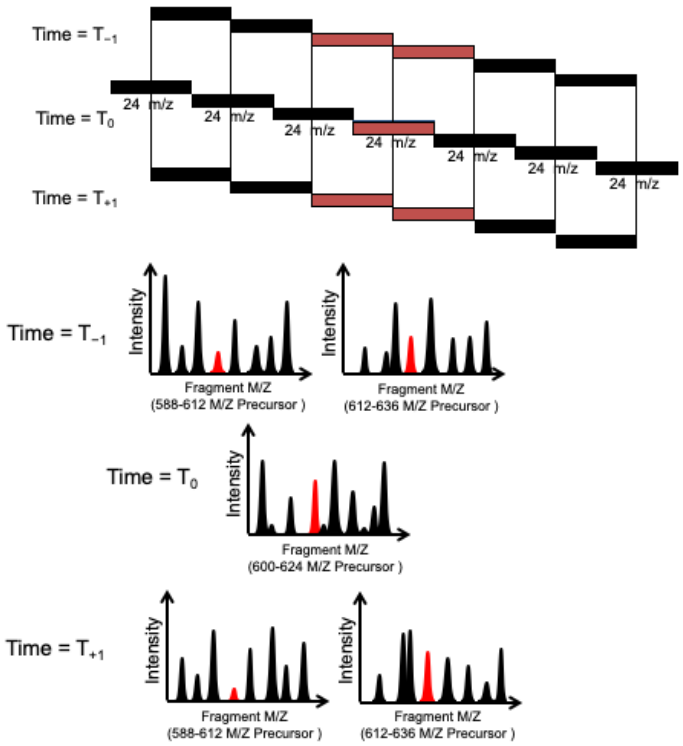


While we can't actually collect spectra faster, we can use under-sampling math to make it seem like we are. If we consider the normal window boundaries, we can stagger every other cycle by 50% using a technique called compressed sensing. Let me give you a feeling for how that works.

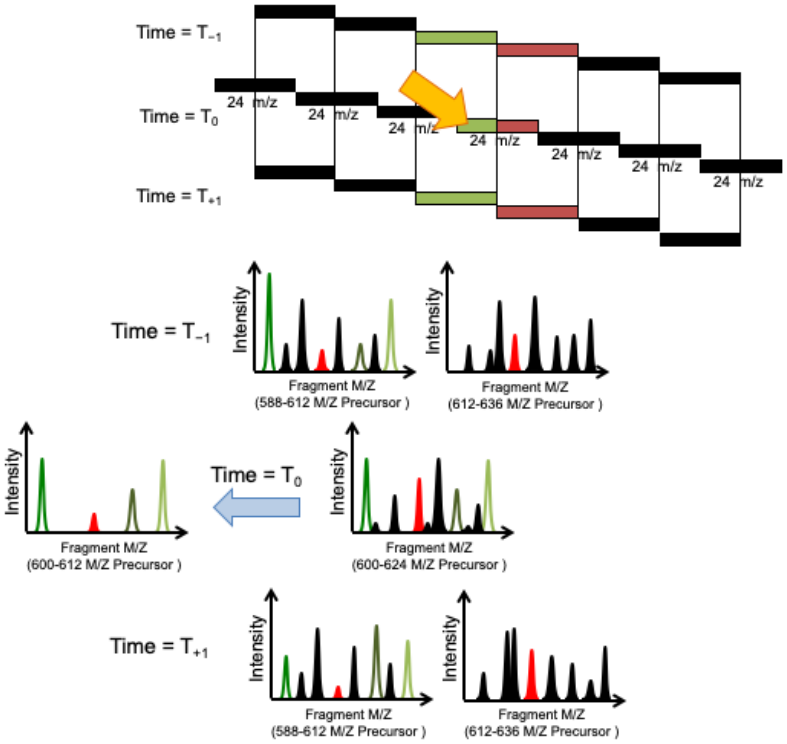




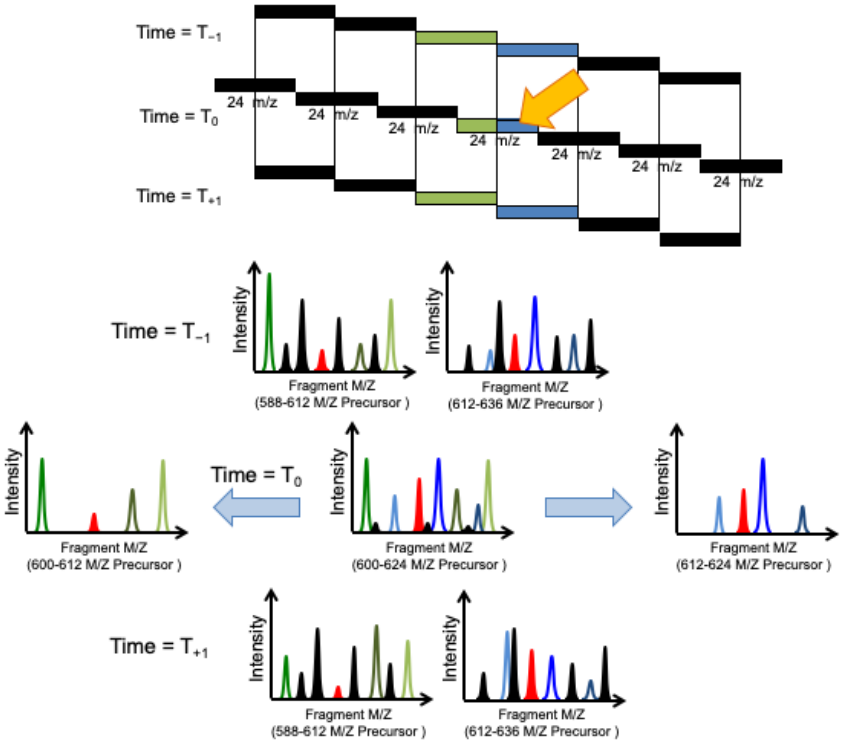
Let's take this 600 to 624 m/z fragment spectrum from a 24 m/z-wide window experiment and consider the previous and next scans.



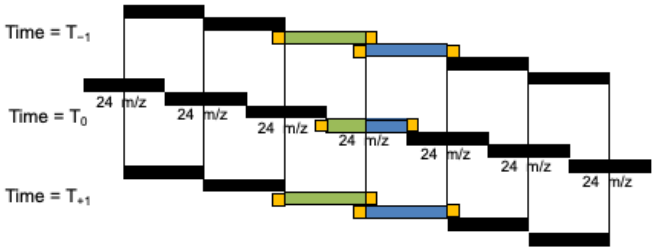
Some ions, such as this red ion, are present in all the scans. We can't really say much about that one.



However, some ions (here in green) are only present in the 588 to 612 m/z previous and next spectra. These can only come from peptides with precursors between 600-612 and we can use linear algebra to create a new spectrum of just these ions.



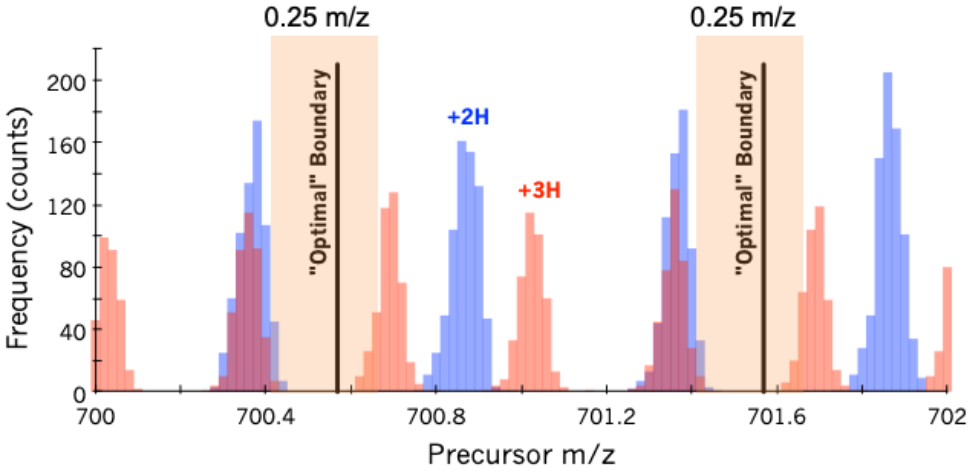
Similarly, there will be other fragment ions that are present only in the 612 to 636 m/z previous and next spectra. These can only come from peptides with 612-624 m/z precursors. Now instead of having 24 m/z windows, after demultiplexing we now have 12 m/z windows, absolutely, 100% for free (with a little extra computational processing). However, this type of math is severely affected by noise. That said, as we discussed earlier, Orbitraps produce relatively little background noise and this approach works remarkably well.



…staggering is incompatible with margins

Now, here's a problem. This staggering approach needs to have fixed window boundaries and is significantly complicated by margins. We can lean on another property of Orbitraps to help with this.

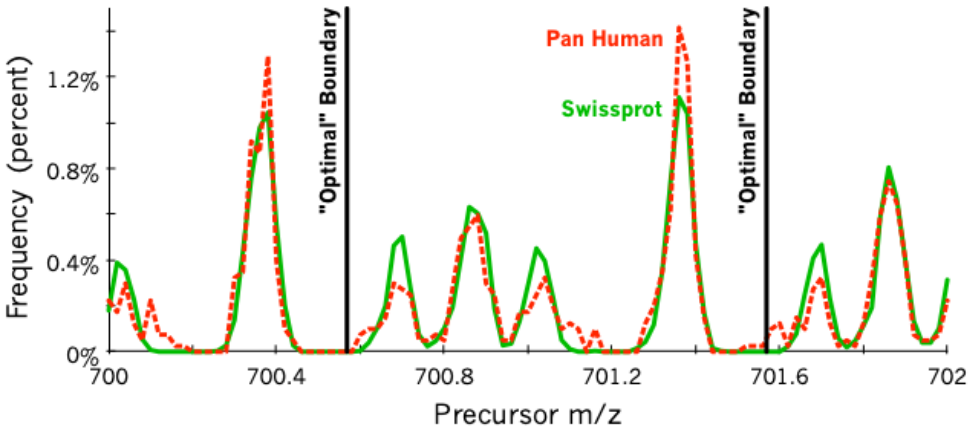
“Forbidden zones” take advantage of m/zs where peptides don’t exist



- Peptides are made of H C N O S

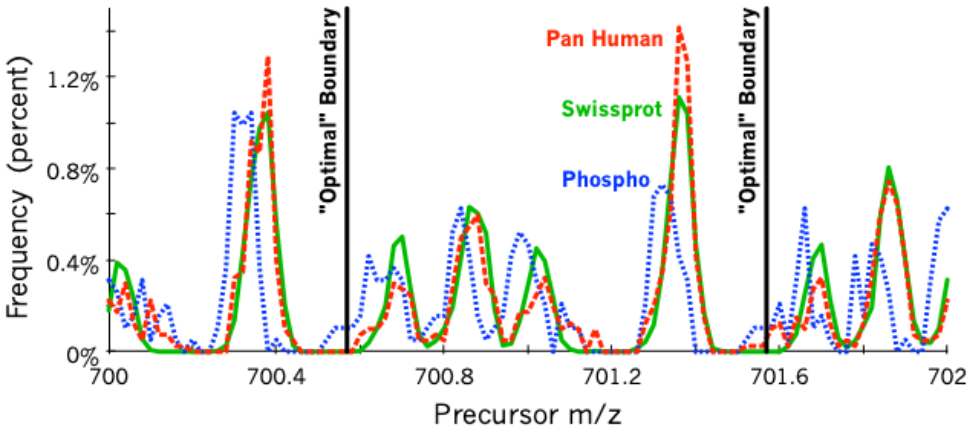
Peptides are all made of the same stuff: carbon, hydrogen, nitrogen, oxygen, and sulfur. There are only so many ways to configure amino acids, and there are sub-m/z regions of the precursor space where you simply cannot have a peptide with that m/z. Here I'm showing that with the distribution of +2 and +3 precursors from tryptic peptides in the Human Swissprot database. These peptides cannot fall in "forbidden zones", which are about a quarter m/z wide. Coincidentally, this also happens to be about twice the edge fall-off of the segmented quadropoles in most current Orbitraps. By placing window boundaries at these fractional masses, we can remove the need for window margins.

“Forbidden zones” take advantage of m/zs where peptides don’t exist



These boundaries hold up looking at actual precursors in the pan-human library

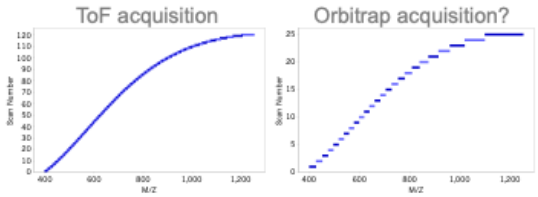
Phosphopeptides have different forbidden zones (-0.18 m/z)



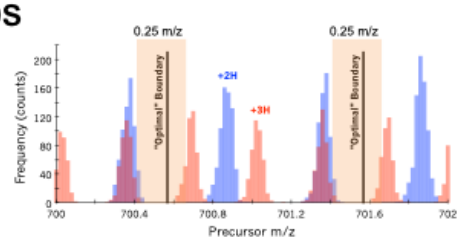
But it should be noted that non-tryptic peptides and peptides with PTMs may have different characteristics. Phosphopeptides, for example, have optimal boundaries that are shifted over by 0.18 m/z.

Contents

- How is DIA different for ToFs and Orbitraps?



- Re-designing DIA to take advantage of Orbitraps
  - Narrow precursor range
  - Stagger windows to under-sample search space
  - Take advantage of “forbidden zones”

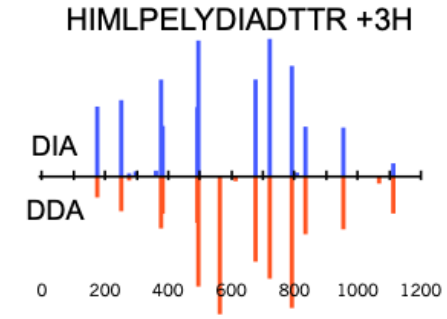


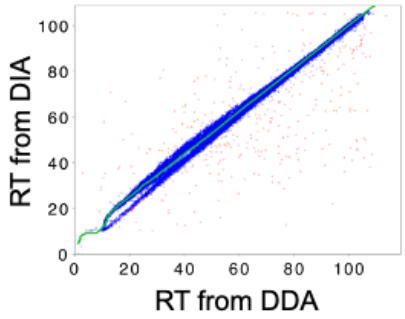
- Generating DIA-only libraries

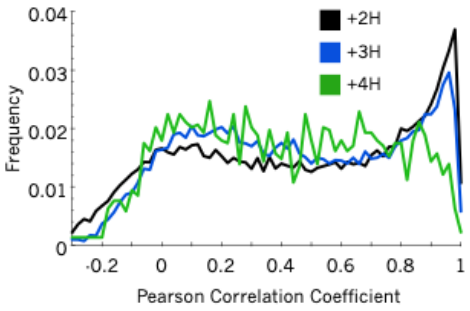
- Resources for DIA best practices

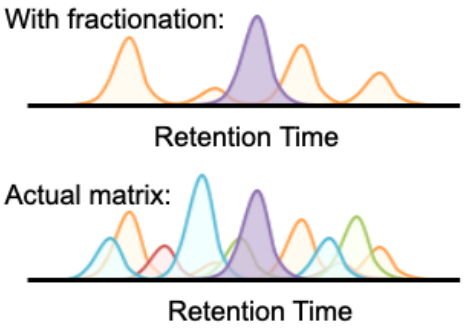
Let’s use the remaining time to talk about building libraries.

Challenges of using DDA libraries



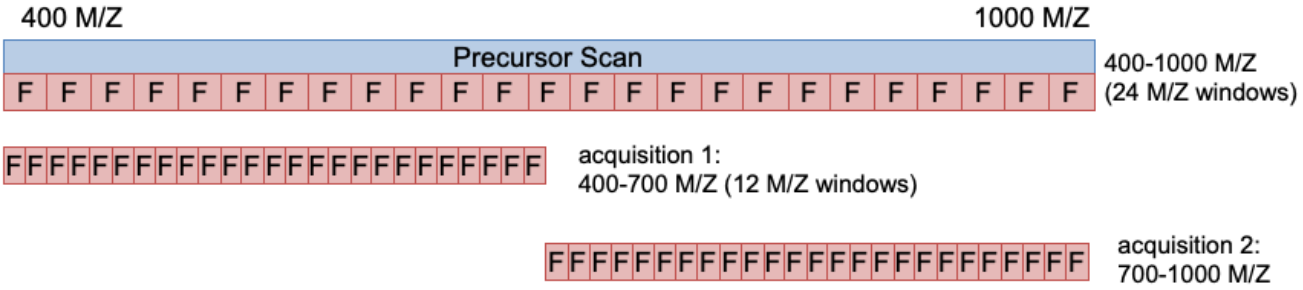






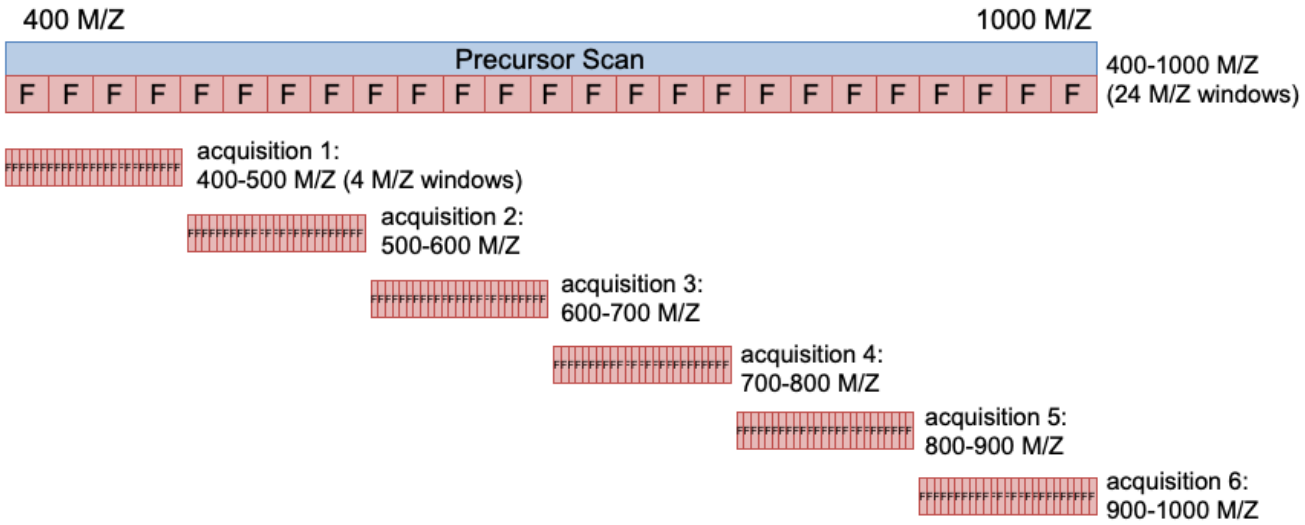
DDA libraries have been intrinsic to the success of DIA, and are used to determine the properties of peptides you want to measure on your instrumentation. However, they are imperfect in several ways. First, these libraries require a significant amount of sample for off-line fractionation, and that can be impractical in some experiments and impossible in others. Second, DDA fragmentation on Orbitraps is charge state optimized, while DIA fragmentation inherently cannot be. This results in small fragmentation variations that can have real effects. Third, because off-line fractionation changes the peptide background, it also subtly changes retention times too. Rather than generate DDA libraries for our DIA experiments, we use an alternate method to generate DIA-only libraries.

Gas phase fractionation



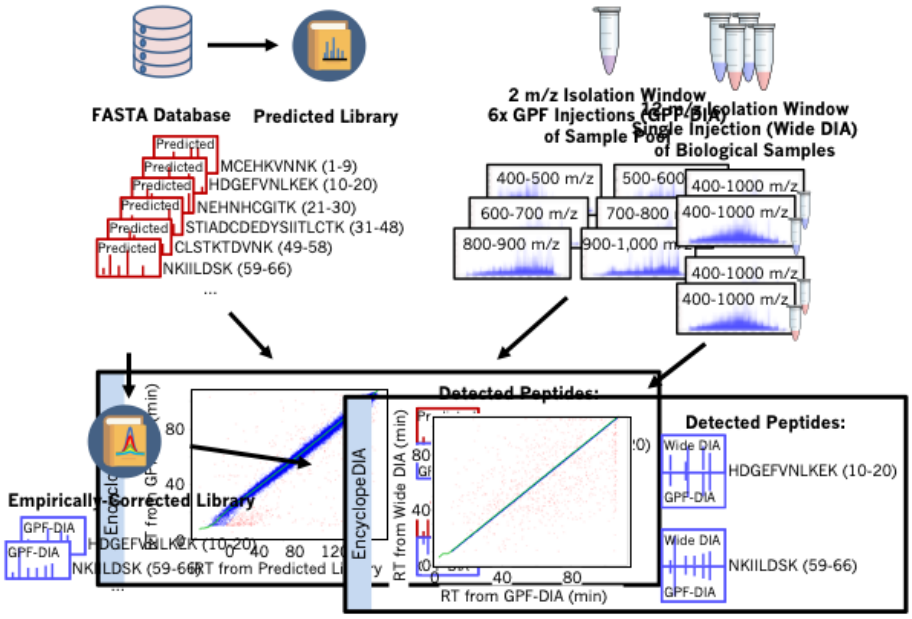
This method relies on gas-phase fractionation, which was originally developed and applied to DIA in the Goodlett lab over ten years ago. The principle is to break up the acquisition of a single sample into multiple runs that cover the same total m/z space with smaller windows. This type of fractionation is interesting because it doesn't involve any off-line work, and it only requires setting up two DIA methods with different windowing schemes.

Gas phase fractionation for library generation



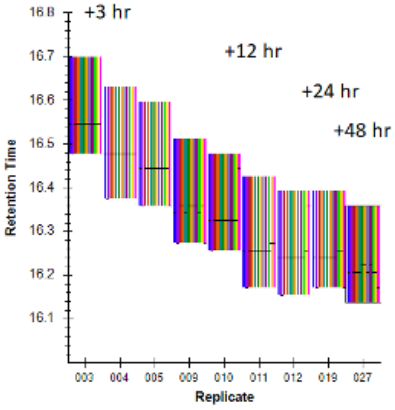
We extend this further out to 6 gas-phase fractions, that each cover 100 m/z. These runs have 4 m/z wide windows, which after staggered demultiplexing, result in 2 m/z wide windows. In a sense, these datasets have the same high quality as targeted PRM experiments, except that they target every peptide from 400-1000 m/z. This uses 6 ug of material, which is also dramatically less than that required for off-line fractionation.

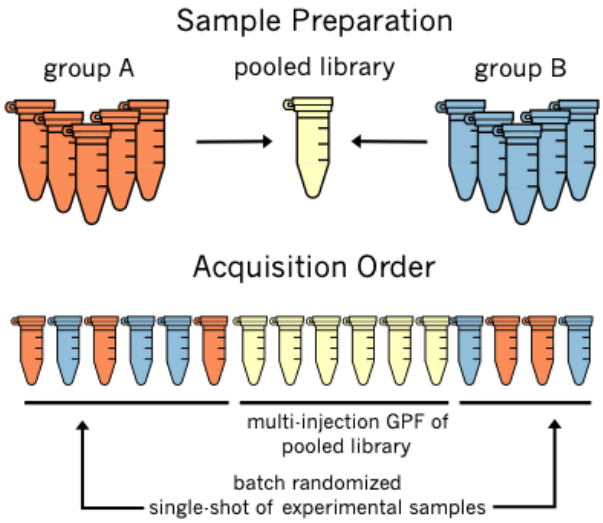
DIA-only libraries with peptide predictions



We collect these 6 injections on a sample pool made from a sub-aliquot of each of our quantitative samples. We then search them using either a FASTA database directly using the PECAN search tool, or with predicted fragmentation patterns using the Prosit tool for every possible tryptic peptide in that database. With effectively 2 m/z wide windows even if the retention times and fragmentation patterns are off, these injections are so clean that it's easy to detect peptides from them. Now that we know the specific DIA-based fragmentation and exact elution characteristics for each peptide in our matrix, it's much easier to look those peptides up in follow-up quantitative DIA experiments.

Order of acquisitions

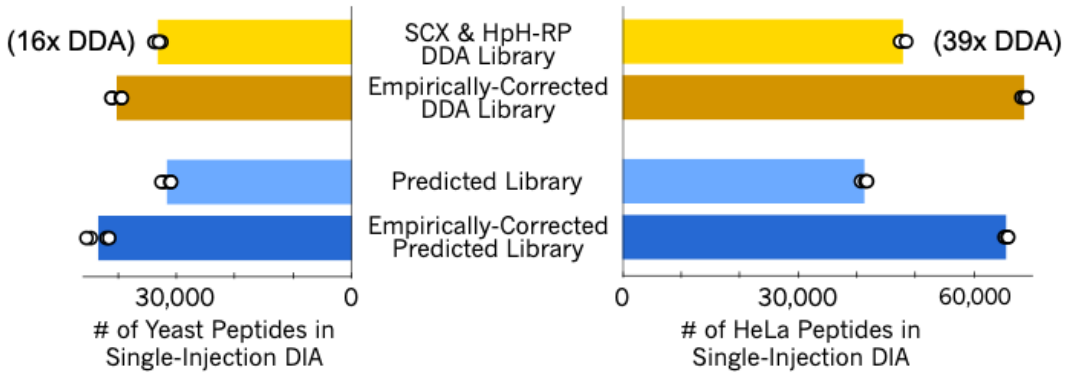




Pino et al, Mol Cell Proteomics 2020

I should note that the 6 gas-phase fractionated injections don't share many peptides, so it's important that there's little retention time deviation between them. We recommend waiting for a column to "break in" and collect the GPF-DIA runs after at least 6 normal injections with the same matrix. While it's difficult to retention time align between the gas-phase fractionated runs to build the library, after it's built, it's easy to align the library to the early normal injections and correct any retention time deviations in those files.

Empirically-correcting libraries dramatically improve peptide detection rates

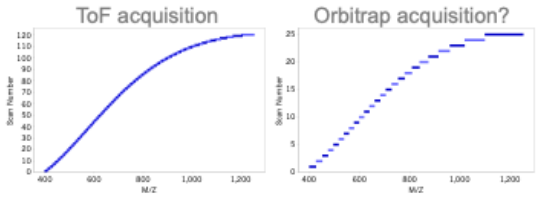


Searle et al, Nat Commun 2020

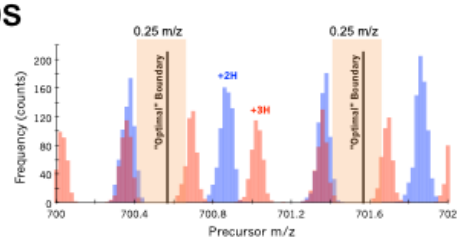
But with these libraries, which again, only require 6 injections and 6 ugs of a sample pool, we can achieve excellent results. Compared to sample-specific DDA libraries using off-line fractionation in yellow, we can use the 6 gas-phase fractionated injections to essentially empirically-correct this library with more exact retention times and fragmentation patterns (in orange) and achieve higher detection rates in both yeast on the left and human cells on the right. Similarly, while Prosit-predicted libraries (in light blue) underperform when compared to fractionated DDA libraries, the performance is very high after correction with the 6 gas-phase fractionated injections.

Contents

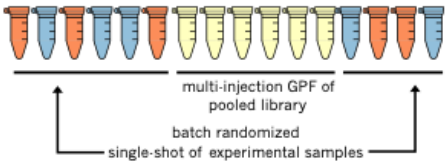
- How is DIA different for ToFs and Orbitraps?



- Re-designing DIA to take advantage of Orbitraps
  - Narrow precursor range
  - Stagger windows to under-sample search space
  - Take advantage of “forbidden zones”



- Generating DIA-only libraries



- Resources for DIA best practices

We’ve developed a bunch of resources to help people start to collect data on Orbitraps using our framework, and I wanted to share them with you.

Resources for DIA best practices

- Perspective preprint: “Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries”
  https://www.mcponline.org/content/early/2020/04/20/mcp.P119.001913
- Quickstart guide for setting up DIA:
  https://bitbucket.org/searleb/encyclopedia/downloads/dia_methods_setup_v1.2.pdf
- Recommended settings for DIA on Orbitraps:
  https://docs.google.com/spreadsheets/d/1A8AQlmLroAkQcAcsiGTNvnGBE2lGpkMwhh0YLTBHXKA
- Download slides and free software for building and searching DIA-only libraries:
  https://bitbucket.org/searleb/encyclopedia

First, much of what we’ve discussed here has just been published in a new preprint in MCP. We’ve built a quickstart guide and a spreadsheet of recommended DIA settings for every QE or tribrid instrument to date. Finally, both these slides and software to implement these methods are freely available at this URL for EncyclopeDIA software tools. Thank you for your attention and I look forward to your questions!

Perspective preprint “Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries”:
https://www.mcponline.org/content/early/2020/04/20/mcp.P119.001913
Quickstart guide for setting up DIA:
https://bitbucket.org/searleb/encyclopedia/downloads/dia_methods_setup_v1.2.pdf
Recommended settings for DIA on Orbitraps:
https://docs.google.com/spreadsheets/d/1A8AQlmLroAkQcAcsiGTNvnGBE2lGpkMwhh0YLTBHXKA
Download slides and free software for building and searching DIA-only libraries:
https://bitbucket.org/searleb/encyclopedia