**Thesaurus: quantifying phosphopeptide positional isomers**

Brian C. Searle, Robert T. Lawrence, Michael J. MacCoss, and Judit Villén*

Department of Genome Sciences, University of Washington, Seattle, WA, USA

*Corresponding author, email: jvillen@uw.edu

**SUPPLEMENTARY NOTE**

**Question: How is reproducibility limited when analyzing phosphoproteomes with DDA?**

The most common mass spectrometric method for large-scale phosphopeptide profiling uses DDA, in which MS survey scans are collected in one to three second intervals and MS/MS peptide sequencing scans are triggered for the top N most abundant ions. A narrow range around each precursor ion is selected for fragmentation, which generates an MS/MS spectrum containing several sequence specific ions. Once an MS/MS scan has been triggered, that precursor mass is placed on an exclusion list, which is designed to limit repeated measurements on the same peptide. This function, called dynamic exclusion, is a key optimization that allows DDA to identify low abundance peptides. However, the tradeoff of deep sampling is that phosphopeptide quantification across multiple runs is often hampered by a lack of consistent detection of low-level phosphorylation events due to stochastic MS/MS sampling. While the accuracy of site-specific phosphorylation calls can be estimated, the general approach is limited because localization is performed using data from a single fragmentation spectrum that is rarely collected near the apex of the peptide chromatographic peak. This can complicate the detection of position-specific fragment ions that do not always rise above the background signal. Even worse, dynamic exclusion prevents detecting multiple positional phosphopeptide isomers by intentionally excluding sampling of other isobaric peptides. Moreover, even if multiple species are detected by DDA, the precursor XIC's that are typically used to quantify these peptides are convoluted and cannot be distinguished.

**Question: How do fragmentation methods affect phosphopeptide fragmentation and influence phosphorylation site localization?**

CID of phosphoserine and phosphothreonine-containing peptides are characterized by a prominent peak attributed the phosphate neutral loss from the precursor (1). The dominance of this fragment is at the expense of sequence informative b- and y-type fragment ions, and therefore has a direct impact in phosphorylation site localization. The intensity of the phosphate neutral loss peak decreases with increased precursor charge, and is significantly less prominent in beam-type fragmentation methods (e.g. HCD) compared with CID (2). Alternatively, analysis of phosphopeptides by ETD does not suffer from the neutral loss problem (3), and typically perform better at size localization due to ETD producing fragment ions that cover most of the phosphopeptide sequence (4, 5). Possibly using DIA with ETD may provide a good approach for analyzing positional isomers (6).
**Question:** With peptide-centric searching I can look for specific positional isomers in PRM/DIA data. Why do I also need to use site localization tools?

While it is possible to use spectrum library search engines to detect phosphopeptides, we do not recommend it without using some sort of localization software. Unlike with spectrum-centric DDA searching where one implicitly competes phosphopeptide localizations against each other (e.g. Mascot Delta Score (5)), with peptide-centric DIA searching one only attempts to find the highest scoring spectrum across retention time as an indication for the presence of a peptide. While this works well for unmodified peptides, shared ions in modified peptides can lead to false identifications. As an illustration, consider the peptide KGSGDpYMPMSPK from Figure 2. Below, Panel A shows the Primary Score (which uses both site-specific and shared fragment ions) trace from a control sample, while Panel B shows the same score trace from an IGF-1 stimulated sample:

![Score traces](image)

Despite that KGSGDpYMPMSPK should be below the level of detection in the control sample, a high score (>14) can be observed at the time point associated with another phosphorylated form, KGSGDYMPMpsPK, at 49.5 minutes (2,970 seconds) because of a large number of shared peaks. Without considering site-specific ions, this score is high enough to be considered a detection at a FDR <0.01. Only when considering an IGF-1 stimulated sample is it possible to determine the true retention time for KGSGDpYMPMSPK at 44 minutes (2,640 seconds).

**Question:** Is it possible to use DDA localization tools for DIA experiments?

Meyer et al (7) have determined that DDA localization tools can be used for DIA experiments if they are deconvoluted using DIA-Umpire. This is successful because DIA-Umpire constructs pseudo-MS/MS spectra that should contain only fragment ions for a peptide precursor m/z. However, most DDA localization tools assume that peptides are isolated independently and are not compatible with DIA without similar deconvolution. For example, Ascore (8), MaxQuant (9), and several other DDA site localization algorithms rely on probabilities generated from a binomial distribution, which enables those algorithms to ask the question “what is the likelihood of drawing a site-specific peak out of a random
Question: Why does Thesaurus out-perform IPF and PIQED at detecting positional isomers?

In Figures 1 and 2, while Thesaurus consistently detected multiple positional isomers, both IPF and PIQED make assumptions that can make it difficult to make these detections. OpenSwath/IPF only considers the top scoring peak group assigned to each isomer. Since positional isomers typically share several fragment ions and OpenSwath scores more intense fragment ions higher, the top scoring peak group is often the same for all isomers (the most intense). Consequently, IPF usually only assigns multiple positional isomers when they either have significantly different sets of fragment ions or have similar abundance. PIQED can consider multiple peak groups but requires that fragment ions fit a precursor peak shape and can miss site-specific ions necessary for localization. Thesaurus uses a peptide-centric approach to identify phosphopeptide sequences but also considers other peak groups determined within a wide retention time window, and thus can resolve chromatographically separated positional isomers that share fragment ions.

Question: Why are there detections of multiple positional isomers at the same retention time?

We find that approximately a 45% of positional isomers elute within five seconds of each other. Our retention time resolution is dependent on cycle time, and as our cycles are around two seconds each, these positional isomers are effectively chromatographically inseparable. In these cases, other localization approaches compete positional isomers found from the same chromatographic peak against each other either and do detect at most only one form. Thesaurus, however, can detect and quantify these using site-specific ions. In many cases we observe a dominant isomer, some of which (for example, Supplementary Figure 4) might be indicative of phosphate rearrangement where the phosphate moves from one site to the other in the gas-phase (10). One way to conclusively differentiate true co-eluting peptides from gas-phase rearrangements is to determine quantitative changes between the forms across multiple conditions (for example, Supplementary Figure 13). In all cases we feel it is critical to quantify co-eluting phosphopeptides using site-specific ions only to avoid contaminated measurements.

Question: What percentage of positional isomer detections are actually gas-phase rearrangements?

Without any additional data, mass spectrometry-based proteomics requires differential quantification to indicate if the co-eluting positional isomers we see are real biological signatures or are gas-phase rearrangements. In an attempt to estimate the effect of gas-phase rearrangements, we used the results of the synthetic peptide experiment shown in Supplementary Figures 3 and 4, collected on a Q-ToF instrument. Here we find that 11 of 240 (4.6%) phosphopeptides produce some signature of gas-phase
In our quantitative MCF-7 experiments, we find a median of 6208 distinct phosphopeptides (sequence+number of modifications) and a median of 916 additional positional isomers, where 84% of those fall with 15 seconds. If we assume that Q-ToF and Q-Exactive gas-phase rearrangement rates are similar, then we can estimate that 6208*0.046=286 positional isomers (or 37% of positional isomer pairs within 15 seconds) may actually be gas-phase rearrangements. While there are several caveats to this estimate, we feel it is a reasonable baseline from which to further study this phenomenon.

SUPPLEMENTARY REFERENCES

1. Beausoleil SA, Jedrychowski M, Schwartz D et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004;101:12130-12135.
2. Moser K, White FM. Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. J Prot Res. 2006;5:98-104.
3. Chi A, Huttenhower C, Geer LY et al. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A. 2007;104:2193-2198.
4. Aguiar M, Haas W, Beausoleil SA, Rush J, Gygi SP. Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. J Prot Res. 2010;9:3103-3107.
5. Savitski MM, Lemeer S, Boesche M et al. Confident phosphorylation site localization using the Mascot Delta Score. Mol Cell Proteomics. 2011;10:M110.003830.
6. Carvalho PC, Han X, Xu T et al. XDIA: improving on the label-free data-independent analysis. Bioinformatics. 2010;26:847-848.
7. Meyer JG, Mukkamalla S, Steen H, Nesvizhskii AI, Gibson BW, Schilling B. PIQED: automated identification and quantification of protein modifications from DIA-MS data. Nat Methods. 2017;14:646-647.
8. Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol. 2006;24:1285-1292.
9. Olsen JV, Blagoev B, Gnad F et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127:635-648.
10. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. Anal Chem. 2008;80:9735-9747.